

## **Section 2**

### **Retrospective Analyses of Spring/Summer Chinook Reviewed in FY 1997**

**with comments by the Scientific Review Panel**

**Reviewer:** Jeremy Collie

**Title of Paper:** *Spawner-Recruit Data for Spring and Summer Chinook Salmon Populations in Idaho, Oregon, and Washington*

**Author:** R. Beamesderfer et al.

Comments: This very useful document is the basis for all the PATH analyses that use stock-recruitment data. The descriptions of each sub-basin give very useful background information on the spawning area of each stock. When surface sediment is expressed as a percentage (page 19) what total is it a percentage of?

a) scientific soundness of the methodology

TABLE 2 is important because it lists the equations used in the run reconstruction. Unfortunately the subscripting is inconsistent, making the calculations difficult to follow. For example  $pE_{ffy}$  gives is the exploitation rate in location  $f$ , fishery type  $j$ , and year  $y$ . What then is  $pE_{Cyn}$  in [14] and [15]? I don't see subscript  $n$  defined anywhere in the table.

What does this statement on page 54 mean? "The brood year age composition was adjusted for the size of the hatchery release, and converted back to return year format?"

b) general suitability of the data for use in the analyses

One point that struck me in reading this report was that there were significant human impacts in many of the sub-basins starting at the turn of the century. Especially in the upper Columbia River, natural spawners are the progeny of hatchery-reared salmon, often from different sub-basins.

Another point that struck me was the lack of age/length samples from earlier years for many of the stocks, especially those from the Columbia River. My perusal gives the following table of when age sampling commenced:

Sub-basin	Year before which average age composition was assumed
Middle Fork Salmon	1960
South Fork Salmon	1960
Imnaha	1961
Grande Ronde	1961
Methow	1985
Entiat	1985
Wenatchee	1973
John Day	1978
Deschutes	1975
Klickitat	1988
Wind	1988

The obvious conclusion is that the run reconstructions for the Snake River stocks are much more reliable than those for the Columbia River.

c) validity of inference and conclusions reached

I agree that the run reconstructions are improvements over simple escapement trend analyses, but the recruitment estimates should not be over-interpreted. The authors note several sources of uncertainty in the run reconstructions, including redd counts, hatchery fractions and harvest rates. Of these, the use of mean age composition to fill in missing years is probably the largest source of error. Without age samples for each year, some recruits are bound to be attributed to the wrong brood year. My concern is that subsequent analyses treat these recruitment estimates as independent observations, when in fact they are not. The analyses most affected by this assumption are those looking for covariability in recruitment (Chapter 2) and the stock-recruitment models. The stock-recruitment models may not be biased, in the sense that a recruit must be attributed to one brood year or another, but variability in recruitment could be attributed to the wrong source, if the variation is actually a measurement error.

In earlier reviews I questioned whether assumptions made in the run reconstructions would influence the correlation structure of the recruitment estimates. I note that many of the multipliers and expansion factors are assumed to be the same for stocks within a sub-basin. These include the sub-basin harvest rate, Bonneville to basin "conversion" (survival) rates, and the Columbia River harvest rate. It follows that recruitment estimates will tend to be more highly correlated *within* sub-basins than *between* sub-basins. The assumptions of common rates are justified and the recruitment estimates are probably the best that can be obtained with the available data. My point is that they are not strictly independent observations and should not be treated as such in subsequent statistical analyses. One must keep in mind that this elaborate house of cards is based on observations of holes in the bottom of streams.

d) suggestions for improvements and extensions to the analytical approaches used

For years in which average age compositions were assumed, it might be possible to adjust the degrees of freedom to account for the lack of independence of the recruitment estimates. Along the lines of Chapter 2, effective degrees of freedom might be the number of years divided by 3, because a recruit of unknown age could have originated from one of three brood years.

Has additional sampling been conducted to test the precision and accuracy of standard stream-sampling techniques? Exhaustive sampling could be conducted in a subset of streams, perhaps with weirs. A student of Carl Walters, Marc Labelle, did this for coho salmon on Vancouver Island. It might then be possible to put confidence intervals on the escapement estimates.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

I see little value in adding index stocks to this set if age composition data are not available for them.

**Reviewer:** Brian Dennis

**Title of Paper:** *Spawner-Recruit Data for Spring and Summer Chinook Salmon Populations in Idaho, Oregon, and Washington*

**Authors:** R. Beamesderfer et al.

Comments:

a) scientific soundness of the methodology

The assembly and reconstruction of the spawner-recruit database used in the PATH analyses is a monumental accomplishment. The report gives a highly polished and complete description of the data, assumptions, and calculations involved. It was highly readable and much appreciated by this reader. Earlier PATH documents had left me quite confused about various aspects of the data.

I do not have much by way of specific comments about the document. The scientific assumptions behind the run reconstructions are out of my areas of expertise (though the discussions of the scientific problems were understandable and well written), and I noted no statistical issues. Peterman (1981) is cited but left out of the reference list.

PATH would benefit from having a greatly condensed, executive-summary style description of the data and reconstructions. Such an account would serve as useful boilerplate for the "data description" sections of the numerous scientific papers that will result from the PATH analyses. Readers would then not have to have the big data document on hand in order to understand the analyses.

b) general suitability of the data for use in the analyses

c) validity of the inferences and conclusions reached

d) suggestions for improvements and extensions to the analytical approaches used.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

**Reviewer:** Saul B. Saila

**Title of Paper:** *Spawner-Recruit Data for Spring and Summer Chinook Salmon Populations in Idaho, Oregon, and Washington*

**Author:** R. Beamesderfer et al.

Comments:

a) scientific soundness of the methodology

It is this reviewer's opinion that the overall methodology is fundamentally sound, and that strong recommendations for dramatic changes in methodology do not seem justified owing to the high variability in the data used in the analysis.

b) general suitability of the data for use in the analysis

It would appear that a thorough search has been made for data used in this study. It is clear that the data are extremely variable, and the results might be even more variable if the estimated errors were propagated through all of the calculations. For example,  $pV$ , the main stem conversion rate ( $pV_{dy}/D_{by} - T_{dy} - E_{cy}$ ) involves terms which are estimated with error, and  $E_{cy}$  contains a sum of terms, which are also estimates. Thus, all sources of error have not been propagated, so the calculated estimates and especially their variation, are considered to be conservative in this report.

c) validity of inferences and conclusions reached

It is difficult to comment on the validity of the inferences and conclusions, because the authors of this document were extremely cautious in making them, and with justification. However, a few comments may seem in order.

Table 5. The coefficient of variation provides a rough measure of relative variation by expressing the ratio of the sample standard deviation as a percentage of the sample mean. It is this reviewer's opinion that for most fishery management-related purposes, the coefficient of variation for estimated stock size (spawners) and recruits should not exceed 25-30 percent. None of the estimates in Table 5 of spawners, recruit, or recruits-per-spawner are within these limits, and only 3 out of 66 are less than 50 percent.

Table 6 illustrates correlations among variables and indicates statistically significant coefficients at the P.05 level or less. In general, this reviewer believes that significant correlations might be further explored for intellectual speculations based on the analysts' experience and judgment. It is noted that this has been done to some extent already. However, it may also be interesting and useful to look at  $r^2$  (the coefficient of determination) which is defined as the proportion of variance in either variable, which is linearly accounted for by the other. The quantity  $r_1^2 - r_2^2$  represents the amount of change in the proportion of variance accounted for. Thus, equal amounts of change can be meaningfully interpreted as equal rates of change in amount of relationship anywhere on the  $r$  scale. An example is provided from Table 7. In examining the  $r$  values for the Bonneville count and the Ice Harbor count, the Bonneville versus Snake River  $r$  values are squared and differenced. In this case, the  $r_1$  and  $r_2$  pairs 0.760, 0.636, and 0.911, 0.711 represent nearly equal differences in the amount of relationship, since in the first pair,  $r_1^2 - r_2^2 = 0.173$ , and in the second pair,  $r_1^2 - r_2^2 = 0.1512$ . The larger  $r_1$  of each pair accounts for about 16 percent more variance than the smaller. Note, however, that the  $r_1$

values are exchanged for the two pairs. That is, in the first pair, Bonneville accounts for more variance, and in the Ice Harbor count, the Snake River accounts for more variance than Bonneville.

- d) suggestions for improvements and extensions to the analytical approaches used

At this point, this reviewer has no suggestions for improvements and extensions. It is suggested that data variability does not permit much hope for improvements.

- e) opportunities for integration of the different component analyses into an adaptive management approach

This reviewer believes that data limitations limit the opportunities for integration of this component into an adaptive management. It is believed that successful adaptive management is not possible with the observed variability in measures of spawner and recruit abundances.

- f) relative priorities for future work on these analyses

It is suggested that future work on these analyses be limited or curtailed due to data limits

**Reviewer:** Carl Walters

**Title of Paper:** *Spawner-Recruit Data for Spring and Summer Chinook Salmon Populations in Idaho, Oregon, and Washington*

**Author:** R. Beamesderfer et al.

I cannot comment on the validity of particular local assumptions used in this synthesis (expansion ratios for redd counts, etc.), but apparently something substantial is missing from the methodology to result in such low correlations as presented between the local spawning estimates and the dam counts. But this concern aside, it is obviously a good idea to have a substantial set of stock time series constructed with a consistent methodology. I do not understand how these data sets relate to the ones used by Rick and others for the retrospective modeling, and would have appreciated a direct comparison of the alternative time series reconstructions. I also do not understand, and would be very concerned about, the methodology used to allocate mainstem total catches to individual stocks; if a lot of fish have been misallocated, this would create all sorts of spurious correlations in the recruitment estimates and could also systematically bias some recruitment estimates (e.g. attributing too many recruits to a stock every year would make that stock look too productive). You need to request a much clearer explanation of this methodology.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Covariability in Abundance Among Index Stocks of Columbia River Spring/Summer Chinook Salmon.*

**Author:** L. Botsford and C. Paulsen

Comments: The new contribution of this chapter is the treatment of intraseries correlation or autocorrelation of the time series.

a) scientific soundness of the methodology

In principle, the level of autocorrelation in the timeseries should be accounted for when assessing the statistical significance of interseries correlation. As the authors point out, high autocorrelation in the timeseries leads to increased occurrence of Type-1 errors, or the acceptance of spurious correlations as significant.

However, the main point of this exercise is to look for spatial patterns of covariability among the index stocks, and the same patterns exist in the correlation coefficients regardless of the significance we assign to them. The methods used in this paper scale down the effective degrees of freedom for each data type by a constant fraction (average  $N^*/\text{average } N$ ). The significance of the correlation coefficients, relative to each other, remains unchanged, although a larger correlation is required for "statistical significance." The authors note on page 9 that they did not account for multiple tests, and therefore that the significance tests can not be interpreted as a true probabilities. Given that the significance levels are only relative to each other, I'm not sure that it makes much difference whether or not they are corrected for autocorrelation.

There are some notation problems in Eq. (1) that make it difficult to understand, and not all the symbols are defined. It seems as if the index  $v$  should be the same as  $j$ .

I agree with the authors that first differences of time series may not be biologically or statistically meaningful (page 4). In the worst case, taking first differences may obscure the pattern of interest.

In Fig. 2 I think that abscissa should be labeled "Correlation Coefficient."

The timeseries  $S_{t+4}/S_t$  is defined as a ratio, not as a residual (see Fig. 3 bottom panels). Since this series is intended as a surrogate for  $\ln(R_{t+4}/S_t)$ , I wonder if it should also be log transformed?

On page 6 the authors note that values of  $\rho$  and  $\rho$  at higher lags are less precise because they are based on fewer data pairs. One way to guard against this is to calculate the autocorrelation coefficients as

At higher lags, there are fewer terms in the numerator, and the autocorrelation coefficients are smaller.

b) general suitability of the data for use in the analyses

I reiterate an earlier concern as to whether the recruitment data from each index stock are truly independent. My understanding of the run reconstructions of Beamesderfer et al. is that there are assumptions about age structure, river passage, etc. that would cause recruitment estimates within a sub-basin to be correlated with each other.



- c) validity of inference and conclusions reached

Some of the interpretation is unclear; for example the last sentence on page 7.

- d) suggestions for improvements and extensions to the analytical approaches used

I tend to favor estimating the effective degrees of freedom on a pairwise basis (e.g. with Eq. 2) notwithstanding the generic approach adopted by the authors. My reasoning is that a pairwise approach would tend to downweight the significance of correlation coefficients of series that were highly autocorrelated. The generic approach downweights the significance of *all* the correlation coefficients, without changing their relative significance.

Another approach would be to test the significance of the correlations with some sort of randomization test, as did Thompson and Page (1989). However, I am not sure how to perform the randomization with autocorrelated timeseries.

- e) opportunities for integration of the different component analyses into an adaptive management approach

These analyses could be useful for prospective models, in that the observed correlation structure in recruitment could be simulated. In simulating the Fraser River sockeye salmon, Collie et al. (1990, CJFAS 47:145-155) used a variance-covariance matrix of recruitment deviations. An alternative approach, used by Deriso, is to partition the recruitment variation into a common year effect and random, uncorrelated deviations.

- f) relative priorities for future work on these analyses

The spatial patterns of variability are interesting but they do not shed much light on the causes of variability. Also, because of concerns about the independence of recruitment estimates, I assign these analyses medium priority for future work.

**Reviewer:** Brian Dennis

**Title of Paper:** *Covariability in Abundance Among Index Stocks of Columbia River Spring/Summer Chinook Salmon*

**Authors:** L. Botsford and C. Paulsen

Comments:

a) scientific soundness of the methodology

1. The authors should clarify the distinction between common trend, covariability (cross-correlation), and synchrony (cross-correlation of noise). These are related, but different, properties of stochastic processes. It is unclear which of these properties is being tested for by the particular formulas listed. The biological interpretations of these different concepts also need discussion.

2. What actual test (test statistic, distribution, decision rule, references) is being used for the significance of the cross-correlations? What time series model assumptions are implicit in the test?

b) general suitability of the data for use in the analyses

The data are suited to the analyses.

c) validity of the inferences and conclusions reached

The possibility that the results of the analyses could be ascribed to the tendency for the stocks from different Columbia River basin watersheds to segregate spatially in the ocean needs to be discussed (Paulsen/Fisher paper).

d) suggestions for improvements and extensions to the analytical approaches used.

Mostly, this reader is confused about aspects of the analytical approaches used (as described in comment a1 above). A main question is, for what kinds of time series are the approaches valid? The series arise from nonlinear processes, and most approaches based on linear processes are questionable. The analysis of the Ricker residuals seems to be the most appropriate use of the analytical methods: the main nonlinearity in the production of the data is already accounted for.

e) opportunities for integration of the different component analyses into an adaptive management approach

"Adaptive management" is a rather ambitious concept for this system, given the general lack of knowledge and the ongoing political interference. This study will help justify a cautious management approach: it provides evidence that suggests the factors imperiling Snake River chinook are peculiar to that watershed.

f) relative priorities for future work on these analyses

Exactly the type of study on which the PATH team should concentrate. The best "decision support" that scientists can provide to "managers" is reliable knowledge.

**Reviewer:** Saul B. Saila

**Title of Paper:** *Covariability in Abundance Among Index Stocks of Columbia River Spring-Summer Chinook Salmon*

**Author:** L. Botsford and C. Paulsen

Comments:

a) scientific soundness of the methodology

This reviewer does not have sufficient background in the analyses of time series to effectively comment on the scientific soundness of the methodology. However, the authors are considered to be competent scientists.

b) general suitability of the data for use in the analyses

It seems somewhat paradoxical that the highest coefficients of variability were found in recruits per spawner in Table 5 of Beamesderfer et al.—*Spawner-Recruit Data for Spring and Summer Chinook Salmon*. How does this reconcile with the statement made here regarding significant covariability? Does taking the logarithm of recruits per spawner influence covariability in recruits per spawner? I realize that this could stabilize variance. Also, it is this reviewer's recollection that the fits of the Ricker stock-recruit function account for far less than 50 percent of the observed variation in the data for virtually all examples. How meaningful are the residuals from such an ill-fitting model which have the highest covariability between sub-basins? Is it reasonable to make strong inferences from such inherently noisy data?

c) validity of inferences and conclusions reached

This review will defer statements regarding the validity of inferences and conclusions until the questions raised in Item b) above are answered

d) suggestions for improvements and extensions to the analytical approaches used

This reviewer is not in a position to suggest improvements and extensions to the analytical methods used.

e) opportunities for integration of the different component analyses into an adaptive management approach

The opportunity for integration of this information into an adaptive management approach seems reasonable.

f) relative priorities for future work on these analyses

The authors suggest testing relatives developed in Chapter 2 through other approaches. This reviewer agrees and recommends moderate priority for this work.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Overwintering Survival of Snake River Spring and Summer Chinook PIT-Tagged Fish*

**Author:** C. Paulsen, R. Hinrichsen, and T. Fisher

Comments: I appreciated the concise presentation of this paper.

a) scientific soundness of the methodology

I am glad to see that my 1996 review stimulated a more in-depth analysis of PIT-tag data. The Poisson model seems to be appropriate, but it could be explained better and the assumptions tested. Eqs. 1 and 2 give the expected number of fish detected. I assume you mean that  $f_j$  are assumed to be known independently so that the  $\ln(Rif_j)$  term is not estimated in the model.

What is the observation model? Was the assumption of a Poisson error distribution tested? In Fig. 2b the distribution of the residuals approaches the normal distribution, but it is still skewed with some outliers. You could try checking the model diagnostics with other assumed error distributions.

As in Table 5, you could test for associations between the habitat cluster categories and the other continuous variables. For example, is habitat category correlated with length to LGR? Such associations between the independent variables could also be tested with GLMs.

b) general suitability of the data for use in the analyses

The authors have assembled a large set of release groups. I don't fully understand how the removal of smolts for barging affects the dependent variables at the lower dams. Were they removed before or after passing through the PIT-tag detectors? Perhaps you could explain the sequence of events in more detail. I assume that removal of large but unknown percentages of smolts for barging contributes to the year and month of release effects in the GLMs. Was the proportion of smolts removed for barging measured?

c) validity of inference and conclusions reached

It is reassuring that in this analysis the outliers had only a modest effect on the results (page 7). This makes the outlier detection/deletion scheme relatively unimportant.

The association of primary interest here is between habitat clusters and recovery proportions. I'm not sure that it makes much difference whether climatic variables are included directly or are just incorporated in the year effects. I guess that using continuous climatic variables gains a few degrees of freedom, but with 1106 release groups degrees of freedom should not be limiting.

d) suggestions for improvements and extensions of the analytical approaches used

I think that the GLMs should be hypothesis driven, instead of fitting a suite of models and selecting one based on goodness of fit or AIC. To test the effect of habitat you need to fit equivalent models with and without habitat clusters and then test for a significant reduction in the deviance with habitat included.

In particular, I think you should test the sub-basin effect I found in my previous review. What is the geographic distribution of the habitat clusters in Table 1? I am not familiar enough with the geography of the Snake River Basin to locate all the release sites by their abbreviated names. It's not that I don't think habitat quality can affect overwintering survival; it's just that the habitat effect could be obscured by other variables.

- e) opportunities for integration of the different component analyses into an adaptive management approach
- f) relative priorities for future work on these analyses

If the tag-recovery data can be expressed as survival rates, it would then be possible to incorporate habitat effects on survival into life-cycle models of Snake River chinook salmon. I therefore give these analyses of PIT-tag data a high priority.

**Reviewer:** Brian Dennis

**Title of Paper:** *Overwintering Survival of Snake River Spring and Summer Chinook PIT-Tagged Fish*

**Authors:** C. Paulsen, R. Hinrichsen, and T. Fisher

Comments:

a) scientific soundness of the methodology

The paper uses generalized linear model analysis to look for patterns in over-wintering survival of PIT-tagged spring/summer chinook of wild origin. The paper is written as an internal memorandum and is hard to follow. The information needs to be organized, filtered, and interpreted for the reader. Nonetheless, the analyses are important, and the methodology is sound.

The main problem with the paper is that the authors do not reach any conclusions. Rather, the paper is just a mass of raw analyses, as yet incomplete. This problem will have to be dealt with before the document is of much use to a reader. I can give a few particular comments on some rough spots:

(p. 2) Identify PTAGIS.

(p. 5) Explain " $\ln(Rif_j)$ " term in Eq. 2 is used as an offset, and the estimated parameter is constrained to equal one."  $R_i$  is known, and  $f_j$  is assumed known (what values?), so what, exactly, is being set equal to one and why?

(p. 5) Why are the wordings of the definitions of " $E(n_{ij})$ " and " $\mu_{ij}$ " different?

The authors posed a few questions for the reviewers, which I will respond to here:

1. "We assume a scaled Poisson distribution ... ." The Anscombe residuals look reasonably normal, at least for model 2B (Fig. 2b) (Let's see these on better quality plots, though.) Thus, the authors can be confident that the scaled Poisson in model 2B captures the essential variability in the data. It is one of the better models on the market for survival count data.
2. "The outlier detection/deletion scheme ... ." Various other outlier detection methods would not likely give results much different from what was reported. However, a distinction should be drawn between "outlier" and "influential observation." An observation can be highly influential to the parameter estimates and results, but not appear as an outlier, and vice versa. The authors should examine a couple of measures of influence, to ensure that the results are not all riding on one or two anomolous observations (see the paper by Hinrichsen et al. in pkg #1; follow their example in analyzing and writing).
3. "Given the focus on habitat ... ." Interactions, it seems to me, are likely to be important in the system, particularly with regard to habitat. Analysis of interactions should be pursued. However, careful attention to the influence measures, and especially, to the BIC model selection index, will help the investigators to retain only those interactions that have genuine regionwide importance.
4. "Because many (if not most) fish ... ." The mark-recapture analyses suggested would be a valuable as a separate check on the results as well as a pilot study on what might be an important way of studying mortality in the future.

b) general suitability of the data for use in the analyses

The data are suited to the analyses.

c) validity of the inferences and conclusions reached

The authors do not really draw any conclusions yet; the analyses are not completed.

d) suggestions for improvements and extensions to the analytical approaches used.

These suggestions are subsumed in my responses to the authors' queries, above.

e) opportunities for integration of the different component analyses into an adaptive management approach

The results of these analyses will be useful in life-cycle simulation models. One could possibly do some experimentation with PIT tag releases in different spatial/habitat locations.

f) relative priorities for future work on these analyses

The habitat question is one of the crucial problems in the survival of the Columbia basin salmon. This study is at the heart of the tasks given to PATH and deserves high priority.

**Reviewer:** Saul B. Saila

**Title of Paper:** *Overwintering Survival of Snake River Spring and Summer Chinook PIT-Tagged Fish*

**Author:** C. Paulsen, R. Hinrichsen, and T. Fisher

Comments:

a) scientific soundness of the methodology

I believe that many of the comments related to the soundness of the methodology made in my review of *Effects of the Ocean and River Environments* ... apply here. In summary, the methodology may be sound, but the statistical assumptions implicit in its valid use have not been adequately tested or stated, in my opinion.

b) general suitability of the data for use in the analyses

It also appears to this reviewer that some of the non-statistical assumptions made in applying the models may not be valid. For example, I question the 1:1 ratio of water spilled to fish spilled as a constant, as well as the assumption that the proportion of fish passing through turbines at PIT-tagging counters was constant among all years.

c) validity of inferences and conclusions reached

I believe the validity of some of the results obtained may also be questioned. Much importance has been placed on the AIC and IBC criteria. It is my opinion that the BIC criterion penalizes model complexity more than the AIC, and perhaps it may be more useful in this analysis as a result. However, it should be clearly recognized that use of AIC and BIC is made under the assumption that a probability process from the family under consideration generates the data. Real world data, such as applied here, never satisfy strict probability premises. The success of one model or another depends much more on the robustness of the model to departures from the assumptions than to theoretical optimality where the assumptions are in fact satisfied. I believe the results reported regarding LGR flow relations are spurious.

Some tentative answers to questions—

- Using a scaled Poisson distribution seems reasonable. However, I don't think that the data will conform precisely to any of the commonly applied distributions. On the other hand, I see no evidence that a systematic testing of various distributional forms was done.
- Outlier detection—I believe the procedure used is a matter of judgment. Plus or minus 3 standard deviations include 99 percent of the observations in a normal distribution.
- I don't believe it is useful to include higher order interactions.
- I think it desirable to do a separate analysis with the slide gate data.

d) suggestions for improvements and extensions to the analytical approaches used

I think that there may be useful alternatives to the method used. Examples include other multivariate techniques such as principal components with instrumental variables (PCAIV), multidimensional scaling (MDS), and possibly neural networks.



- e) opportunities for integration of the different component analyses into an adaptive management approach

I do not see much opportunity for integration at this point.

- f) relative priorities for future work on these analyses

I would place a relatively high priority on reanalysis of this data set by alternate methods.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Chapter 4 Update: Effects of Climate and Land Use on Index Stock Recruitment*

**Author:** C. Paulsen, R. Hinrichsen, and T. Fisher

Comments: Overall, I found this update to be a much-improved version of Chapter 4.

a) scientific soundness of the methodology

I support the methodological changes that were made, namely: inclusion of SST as a climate variable, summation of recruits over all return years and the efforts to correct for autocorrelation. I think that these changes improve the quality of the analyses.

Anderson & Wilen (1985, CJFAS 42:459-467) used a three-step regression approach to analyze cross-sectional data on coho stocks. This method corrected the regression estimates for heteroskedasticity and correlation between stocks, and autocorrelation of residuals within stocks. These corrections may be appropriate in this study.

b) general suitability of the data for use in the analyses

Were the climate variables smoothed for the regression model or just for presentation in Figs. 1-4? I have not yet seen the ocean distribution analysis of CWT data, but find it intriguing that the different stocks may experience different ocean conditions.

How exactly were missing values handled, especially given that the logging data were of unequal length in the 8 sub-basins?

c) validity of inference and conclusions reached

The question implied by these analyses is whether there are environmental factors unique to the Snake River stocks that can explain their lower survival rates since the period of mainstem dam construction? In Chapter 5 Deriso et al. fit a model that contained shared year effects of all stocks (\*) and a  $\gamma$  parameter that was interpreted as passage mortality for Snake River stocks. Model 5 differs from Deriso's in that there are no year effects included and it is not clear that passage mortality has been parameterized in the same way. I guess the  $\gamma$ 's are the year effects for Snake River and Mid-Columbia stocks.

What quantity is being plotted in Figs. 6A and 6B for Model 4? The text states only that they are "differences in upstream/downstream recruitment from the climate models."

To reiterate, the important question here is not so much which model explains more of the variability, but to what independent factor is that variability attributed?

d) suggestions for improvements and extensions to the analytical approaches used

I would expect the effects of logging to persist for longer than one year. I suggest using a cumulative index of logging with earlier years discounted by 1 minus the inverse of the forest regrowth time:

$$CUMTIMBER_t = \sum_{i=1}^t PERTIMBER_i \left(1 - \frac{1}{\tau}\right)^{t-i}$$

where  $\vartheta$  is the time for forest regrowth

- e) opportunities for integration of the different component analyses into an adaptive management approach
- f) relative priorities for future work on these analyses

**Reviewer:** Brian Dennis

**Title of Paper:** *Chapter 4 Update: Effects of Climate and Land Use on Index Stock Recruitment*

**Authors:** C. Paulsen, R. Hinrichsen, and T. Fisher

Comments:

a) scientific soundness of the methodology

The authors describe revisions in the analyses of the PATH Chapter 4. First, new ocean indices are used in place of the NPI index. Second, land use data on fire, grazing, and logging are incorporated. Third, the Ricker-style modeling approach of Chapter 5 is used. The new ocean indices appear to be superior and the analyses are on track.

One concern (p. 4) that the authors note is the possibility that the trends in climate variables and the trends in index stock survival are merely coincidental. This is, of course, the drawback of all observational studies: coincidental relationships cannot be ruled out by the data alone, and conclusions typically need to be supported by outside evidence. The conclusions of this study will depend on the concordance of the results with other knowledge about the system and will have to be carefully phrased and discussed.

A concern I wish to note is about the degrees-of-freedom adjustment (p. 5-6) used in the t-tests. I have not understood Botsford and Paulsen's discussion of this adjustment. It goes against all statistical instincts. In normal linear models, in which the observations on the "dependent" variable are independent, various statistics have t, chi-square, or F distributions under appropriate null hypotheses. When observations are not independent, such as in time series models, the problem is not that the degrees of freedom in the t, chi-square, and F distributions are wrong. The problem is that the statistics no longer have t, chi-square, or F distributions. Adjusting df's is a formula-based kludge that may or may not work depending on circumstances. It is not clear whether statistical theory would justify this fix or not (would help to give readers a reference list). A solution based on statistical principles is to use the data to estimate the appropriate distribution of the test statistic (bootstrapping).

In any event, a focus on hypothesis testing of individual coefficients is probably not necessary. The study is a model selection problem, and the interpretation of results should center on the BIC values of the various models fitted. Models with autocorrelated noise, for instance, might be included in the gallery of models under consideration.

b) general suitability of the data for use in the analyses

The data are suitable for the analyses.

c) validity of the inferences and conclusions reached

See a) above.

d) suggestions for improvements and extensions to the analytical approaches used.

In a) above.

e) opportunities for integration of the different component analyses into an adaptive management approach

The approach is compatible with the planned prospective analyses.

f) relative priorities for future work on these analyses

The study is central to the PATH investigations.

**Reviewer:** Saul B. Saila

**Title of Paper:** *Chapter 4 Update: Effects of Climate and Land Use on Index Stock Recruitment*

**Author:** C. Paulsen, R. Hinrichsen, and T. Fisher

Comments:

a) scientific soundness of the methodology

In general, the precautionary comments and suggestions related to the scientific soundness of the methodology made on Package #1 documents which involved the use of generalized linear models (GLMs) apply here. However, it is recognized that substantial efforts were made here to estimate the effects of autocorrelation in independent variables, and testing for the significance of estimated parameters was also done. In spite of this, I have some reservations regarding the model in terms of overparameterization and non-linearity in some parameters. In my opinion, there is no "magic bullet" for addressing these concerns. However, I believe that it is useful and informative to utilize (if possible) several independent approaches to this type of problem. A number of multivariate alternatives, such as principal components with instrumental variables, might be considered. Multidimensional scaling would handle non-linearities. I believe that the literature dealing with using multivariate methods for similar problem types has not been adequately reviewed. A few examples of possibly useful literature include Pech and Laloe (1997), Walker and Saila (1984), and Pearcy et al. (1996). Principal components analyses and other techniques were used by Jamir et al. (1994) to develop predictors of coho salmon survival. These included the Baker upwelling index, an expression of Columbia River influences and indicators of El Niño southern oscillation conditions. Question—Why haven't examples, such as the above, been more effectively considered in the PATH analyses? In summary, I believe the comments and suggestions made here apply to other applications of GLM in the PATH documents.

b) general suitability of the data for use in the analyses

As indicated by the authors, there are lots of possible variables that might be used. It seems to me that a more careful screening of somewhat similar applications with other salmonid species might be helpful in obtaining parsimony in variables. See comments in a) above.

c) validity of inferences and conclusions reached

At this point, I suggest that the validity of inferences and conclusions from this type of analysis should be treated with caution. Spurious correlations have been mentioned briefly in this report and in some previous applications of GLM in Package #1. It has been recognized for some time that spurious correlations are a very common statistical pitfall. See, for example, Scott (1979). It should be kept in mind that statisticians agree that the factual interpretation of any statistical functions and their testing for possible causal relations, are outside the scope of statistical methodology. Even when stochastic dependence seems certain, a functional relation—such as the number of storks and newborn babies in Sweden—says nothing about a causal relation. The issue of spurious correlation is addressed in a general way—not as a specific criticism of this chapter.

- d) suggestions for improvements and extensions to the analytical approaches used

Some suggestions for improvements and extensions have been made in the previous sections.

- e) opportunities for integration of the different component analyses into an adaptive management approach

I think it is appropriate to refrain from integration of the current analysis into an adaptive framework, until more confidence in the results is obtained.

- f) relative priorities for future work on these analyses

There is scope for further work, and this should be given moderate priority.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Chapter 10 - Version 2.0: Freshwater Spawning and Rearing Habitat*

**Author:** M. Jones

Comments:

a) scientific soundness of the methodology

I find it convenient to consider the Ricker a parameter as related to habitat quality (per capita productivity at low stock sizes) and the Ricker b parameter related to habitat quantity. Given the qualifications mentioned on page 1, how can habitat quantity be measured so that it can be linked to the Ricker b value? One possibility that is mentioned on page 12 is estimates of smolt production capacity on a reach-by-reach basis.

b) general suitability of the data for use in the analyses

c) validity of inference and conclusions reached

Section 10.2 provides indirect evidence for the effects of habitat change on chinook salmon productivity. However there are no data that could be used to quantify the increase in productivity that would be expected from habitat improvements.

Retrospective analyses indicated that habitat quantity is not the main factor presently limiting chinook salmon survival (e.g. Chapter 9). It makes sense that when the runs are depressed, freshwater rearing habitat is not the limiting factor. It follows that modifications that increase habitat quantity (e.g. the amount of pool habitat, or the length of stream habitat with a suitable temperature range) may not have discernible short-term effects on chinook production. Habitat quantity could become a limiting factor when the runs have rebuilt and more spawners are required to compensate for other sources of mortality.

d) suggestions for improvements and extensions to the analytical approaches used

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

Given the stated limitations in measuring habitat attributes and in relating them to chinook salmon productivity, I give these analyses medium priority for future work.

**Reviewer:** Brian Dennis

**Title of Paper:** *Chapter 10 - Version 2.0: Freshwater Spawning and Rearing Habitat*

**Authors:** M. Jones

Comments:

a) scientific soundness of the methodology

The first part of this paper is a useful review of existing evidence regarding the effects of spawning and rearing habitats on Columbia River spring/summer chinook salmon stocks. The paper then describes analyses that are planned or underway, and briefly discusses the implications of the work for the prospective analysis. The paper is not yet finished.

Four analysis projects are described:

1. Eastside Assessment (EA). A multivariate analysis of chinook stocks and the landscape features in the EA database indicated a possible relationship between road density and stock status in the Columbia basin. The result is intriguing and, needless to say, important to develop and communicate completely.
2. Index stock analysis. The planned analyses involve looking at how Ricker model parameters vary with EA habitat quality ratings. The analyses represent an excellent direction for PATH to pursue, particularly for testing a priori hypotheses about habitat effects in the system.
3. Idaho parr density analysis. The planned analyses will use multivariate methods to search for relationships between habitat/land use features and parr densities in Idaho streams. The analyses would appear similar to 1 above and should yield interesting results.
4. PIT tag data (this is the analysis described by Paulsen, Fisher, and Hinrichsen in Package #1).

b) general suitability of the data for use in the analyses

The analyses in all cases appear to be suitable for the data.

c) validity of the inferences and conclusions reached

Of course, the data sets are observational in nature, but the facts that the data are still being assembled and that specific hypotheses of interest are spelled out in advance serve to strengthen the inferences planned.

d) suggestions for improvements and extensions to the analytical approaches used.

Make sure a priori hypotheses are spelled out explicitly, so that the analyses are not just fishing expeditions.

e) opportunities for integration of the different component analyses into an adaptive management approach

Ties into the prospective analyses.

f) relative priorities for future work on these analyses

It is important to get this finished and written up!



**Reviewer:** Saul B. Saila

**Title of Paper:** *Chapter 10 - Version 2.0: Freshwater Spawning and Rearing Habitat*

**Author:** M. Jones

Comments:

a) scientific soundness of the methodology

It is difficult for this reviewer to judge the scientific soundness of the methodology for several reasons.

- Mention was made of a "multivariate (classification tree) analysis" (page 10, middle of page and page 11, first paragraph). No references were provided concerning the specific methodology, and I am not familiar with the phrase (multivariate classification tree). This technique should be clearly explained and cited. Table 10.1 is virtually useless without further understanding of its derivation and application.
- Multivariate methods were mentioned with no further detail (10.3.3) Idaho parr density analysis. What specific methods are to be employed and why? I have some suggestions regarding specific approaches to these data and perhaps other data in the PATH analysis.
  - 1) I mentioned principal component analysis with instrumental variables (PCAIV) as a possible alternative or addition to the generalized linear model (GLM) procedure discussed in PATH Package #1. I believe PCAIV (Pech and Laloe 1997) should be investigated as a possible specific method of potential value for the Idaho parr density analysis.
  - 2) Meta-analysis is a term widely used in biomedical statistics and to an increasing amount in ecology. See, for example, Warwick and Clarke (1993) for an application. Meta-analysis refers to the combined analysis for a series of case studies that in them are of limited value but in combination provide a more global insight to the problems under consideration. A book by Eddy, Hasselblad, and Shachter (1992) describes a form of meta-analysis that is described as the confidence profile method. A demonstration diskette of a software program FAST-PRO is available with the book, and a professional version can be purchased. Although the examples used in the book are almost exclusively biomedical, it is my opinion that the technique can also be of potential value for the problem at hand. At the least, it deserves some exploration for possible use.

b) general suitability of the data for use in the analysis

It is also difficult to assess the suitability of data for use in the analyses without seeing the analyses. However, I believe that a significant amount of effort has been applied to obtaining relevant literature and data. My only suggestion regarding data is that literature and data from other regions might be used in a type of meta-analysis to provide more information on effects of habitat changes on salmonids, for example.

c) validity of inferences and conclusions reached

I believe the inferences drawn from the data are valid, but I think a meta-analysis type of approach similar to that indicated above could expand this.

d) suggestions for improvements and extensions to the analytical approaches used

I have made several suggestions for improvements and extensions in the prior comment sections. In general, I believe that a broader search for comparative literature and data from other regions and other salmonid species is appropriate and that this information should be addressed in the context of a meta-analysis. I believe the confidence profile method is one possible approach to meta-analysis of this type of data.

- e) opportunities for integration of the different component analyses into an adaptive management approach

It seems premature to speculate on opportunities for integration without a completed report.

- f) relative priorities for future work on these analyses

I believe the relative priorities for completion of this work should be very high. I believe that priorities for future work should be deferred until this work is completed.

**Reviewer:** Carl Walters

**Title of Paper:** *Chapter 10 - Version 2.0: Freshwater Spawning and Rearing Habitat*

**Author:** M. Jones

In my view, talking about habitat change and how it ought to affect fish is a total waste of time. The proof is in the pudding, i.e. direct demonstration of whether or not there have been changes in f.w. survival (rearing or smolt ratios to estimated egg dep.) or in  $\ln(R/S)$  in situations where the  $\ln(R/S)$  changes are demonstrably not confounded with (or highly correlated with) other known factors that can influence overall survival. I applaud the intention to do such analyses, and predict that they will show no consistent effects anywhere in the system. I expect the outcome to be much like has been obtained from looking for R/S correlations with area logged, there's nobody home and no way to decide whether this represents lack of effect or inadequate index to use as a statistical independent variable.

For prospective planning, perhaps you should include habitat restoration/improvement measures as a basic experimental treatment option along with transportation and drawdown treatments. A lot easier to design and carry out, since can obtain good replication and treatment-reference comparisons within several sub-basins. Treating habitat management as part of the overall adaptive management plan would make that plan a nested experimental design (habitat treatments within watersheds, within temporal treatments applied to the whole system) such that the habitat treatment effects should not be confounded with whole system effects.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Effects of Ocean and River Environments on the Survival of Snake River Stream-Type Chinook Salmon*

**Author:** R. Hinrichsen, J. Anderson, G. Matthews, and C. Ebbesmeyer

Comments:

a) scientific soundness of the methodology

The treatment of points with undue influence is questionable. It is difficult to justify removing data points just because they have a large influence on the fitted model. Perhaps the data could be scrutinized for an explanation of why, for example, tag group #150 had such a large return. Perhaps this is due to an observation error (greater effort in detecting tags) or some other covariate that affects survival. Unless there is some independent grounds for downweighting or removing data points, I tend to favor Model D over Model C.

I don't think that the three age groups returning from the same release should be treated as independent observations. I made this point in earlier PATH reviews, and have not heard a compelling argument for doing so. Treating each age group as an observation inflates the degrees of freedom of the model. Clearly the returns at age of a single release are not independent because all the ages experience the same conditions until their return year. Most of the variation in survival occurs in the early years (downstream migration and first-year of marine life) when the different age groups experience the same conditions. As the authors point out on page 12, it is impossible to separate variations in survival from variations in maturity. Yet variable age of return is thought to be more due to changes in maturation than to ocean mortality (page 40). Therefore there is no point in trying to model the survival of each age group separately.

Justify the choice of the scaled Poisson distribution for these observations and write out the full model in the methods section. Otherwise it is difficult to understand whether the response variable is the recovery proportion,  $p_{ik}$  (p. 11),  $\log(p_{ik})$  (p.36) or  $\text{logit}(p_{ik})$  (page 29). The fact that all the citations are not included in the REFERENCES section doesn't help.

b) general suitability of the data for use in the analyses

The assumption of a constant collection efficiency ( $f=0.4$ ) is quite strong and needs to be examined and justified. It seems unlikely that  $f$  has been constant and I wonder if efforts have been made to measure it directly? Variations in  $f$  could introduce spurious year and/or period effects, and could account for anomalously high or low survival rates. My main point is that it is not worth running too far with GLM if two of its main assumptions (independence of data, and response measured without error) are conspicuously violated.

c) validity of inference and conclusions reached

I agree with conclusions 2, 3, 6 and 7 on page 45.

The observation that age of return explained most of the model deviance (1) and that the age distribution of returning fish varied between years is a trivial result. This variability needs to be factored out of the response to examine the main effects of interest.

Why would you expect flow to affect the survival of smolts that are transported (page 43)? Transported smolts would seem to be removed from the direct effects of river flow; the effects of river flow on

estuarine conditions could potentially influence their survival after their release. To test the direct effects of river flow on smolt survival, you would need to examine in-river migrants.

These results do have interesting implications for hatchery practices. If large numbers of hatchery-produced smolts continue to be transported around the mainstream dams, it may make sense to time the smolt releases to near-shore conditions and not to flow conditions in the Snake and Columbia Rivers.

d) suggestions for improvements and extensions of the analytical approaches used

I suggest summing the recoveries over all return ages and using this as the response variable. The proportion returning at age 4 could be included as an independent variable in the GLM if it explains some of the deviance in recovery rates. This would factor out the effect of age structure while maintaining the true number of observations (50 tag groups) and degrees of freedom.

e) opportunities for integration of the different component analyses into an adaptive management approach

I think it is premature to incorporate the selected statistical model into CRiSP and other life-cycle models. I thought that CRiSP was a dam passage model and would not pertain to smolt-adult survival of transported smolts?

f) relative priorities for future work on these analyses

I agree that if smolt transportation is continued, it should be monitored. However, this type of tagging study does not address the more fundamental question of whether transportation is effective in increasing smolt-to-adult survival rates. I therefore give a low priority to future analyses of this type.

**Reviewer:** Brian Dennis

**Title of Paper:** *Effects of Ocean and River Environments on the Survival of Snake River Stream-Type Chinook Salmon*

**Authors:** R. Hinrichsen, J. Anderson, G. Matthews, and C. Ebbesmeyer

Comments:

a) scientific soundness of the methodology

The paper provides a comprehensive analysis of survival during the oceanic portion of the Snake River spring chinook life cycle. The analysis uses state-of-the-art generalized linear statistical modeling techniques. The methodology is sound and well described.

b) general suitability of the data for use in the analyses

The data are well suited to the analyses.

c) validity of the inferences and conclusions reached

The conclusions are clear and well supported by the analyses. The use of model diagnostics is outstanding.

d) suggestions for improvements and extensions to the analytical approaches used.

Proper use of the AIC model selection index involves fitting all possible models within a family of models, rather than fitting stepwise. There is the possibility that a good model is being missed among all of the covariates by the stepwise model selection approach.

e) opportunities for integration of the different component analyses into an adaptive management approach

1. Hatchery releases are implicated in this study as a possible factor negatively influencing survival. Hatchery releases might be a politically acceptable arena for long-term experimentation (unlike drawdowns); cleverly designed release experiments that react to prevailing river & oceanic conditions might be able to yield valuable information.

2. The results of this study can be used directly in life cycle simulation models.

f) relative priorities for future work on these analyses

The analyses deserve high priority. Although the analyses are in a nearly finished state, reporting the results to the scientific community should proceed without delay. Also, the analyses (and the production of the underlying data!) should continue on an ongoing basis, for monitoring purposes and for studying the system under a wider variety of environmental conditions.

**Reviewer:** Saul B. Saila

**Title of Paper:** *Effects of Ocean and River Environments on the Survival of Snake River Stream-Type Chinook Salmon*

**Author:** R. Hinrichsen, J. Anderson, G. Matthews, and C. Ebbesmeyer

Comments:

a) scientific soundness of the methodology

The class of generalized linear models (GLMs) includes multiple regression, log-linear, logistic, proportional hazards, and probit models. These models have the property that the observations to be accounted for by some explanatory variables  $X_1, X_2, X_p$  take the following form:

$$Y + f(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) + e,$$

where  $f(z)$  is one of several possible functions, and  $e$  is an error term from one of several possible distributions. It is believed that generalized linear models may provide a scientifically sound approach to the selection of variables and their evaluation for predictive purposes. However, there are several caveats regarding this approach which should be carefully considered. Some of these follow:

- An important characteristic of GLMs is that their use implies independent (or at least uncorrelated) observations. Data exhibiting significant autocorrelations of time series or spatial processes should be excluded from consideration according to McCullagh and Nelder (1989). Have autocorrelated data been adequately eliminated from this application? The assumption of independence is a well-known characteristic of classic regression models. I believe it carries over completely to the wider class of GLMs.
- The stepwise regression procedure, which involves both backward elimination and forward selection until both fail to change the model, ignores any relationships that may exist among the covariates.
- Developing an effective GLM model involves the choice of a suitable link function and error distribution, as well as finding a parsimonious model. This latter issue implies excluding covariates having no detectable effect on the response variable. Has this been accomplished effectively?
- Use of deviances as a generalized measure of discrepancy in the form of analysis of deviance tables is much more complicated than a similar analysis of variance table due to the non-orthogonality of the terms in spite of statements suggesting they are analogous. The "bottom line" is that it seems highly unlikely that a single model is clearly the "best" when several others are fairly close to it in terms of goodness of fit.

b) general suitability of the data for use in the analysis

The response variable data (survival index) used was obviously limited by the duration and nature of the marking, tagging, and recovery program. Inferences from these data should be limited and very carefully stated. In my opinion, the fact that all smolts marked and tagged were barged several hundred kilometers to below the last dam (most of the total outmigration distance) is an important consideration. The addition of information from this analysis to the CRISP is not thought to be useful

or even appropriate at this point. The treatment of fish trap efficiency, as a constant of 0.4 is open to question. I believe that trap efficiency probably varies across tagged groups. This should be tested

c) validity of inferences and conclusions reached

I have some questions regarding the validity of inferences and conclusions. They are related to some of the statements regarding the soundness of the scientific methodology suitability of the response variable and some of the covariate data.

The following are some specifics believed to be related to the validity of inference and conclusions.

- Since the fraction of the total outmigration distance that these smolts were barged was the greater part of the total outmigration distance, the validity of inferences that higher survival rates later in the season were not attributable to increased flow is questioned. Furthermore, the bulk of evidence to date indicates a strong positive (but non-linear) relationship between increased flow and survival. This should be given more weight than these results from the GLM model. I believe that the inferences made in the last paragraph (page 41) are not justified, because the data do not permit an adequate test of the relation between survival and flow. Please explain the statement ending with "and presumes 15% of the overall smolt migration were wild fish." Also explain the way the wild numbers for each date were derived

d) suggestions for improvements and extensions to the analytical approaches used

I believe that alternative methods are possible and that they should be examined. For example, use of principal component analysis with instrumental variables (PCAIV) should be explored. See, for example, Pech and Laloe (1997).

e) opportunities for integration of the different component analyses into an adaptive management approach

I do not at this point encourage integration of this material into any adaptive management approach.

f) relative priorities for future work on these analyses

I believe future work should carefully assess alternative methods for analyzing these data.



**Reviewer:** Jeremy Collie

**Title of Paper:** *Update on Ocean Distribution of Coded Wire Tagged Spring/Summer Chinook*

**Author:** C. Paulsen and T. Fisher

Comments:

a) scientific soundness of the methodology

The GLIM seems to be fairly straightforward and is the same model that has been used in previous chapters, to analyze PIT tag returns, etc.

b) general suitability of the data for use in the analyses

In response to reviewers' comments, the authors have added some tag groups and restricted the analysis to release years that are the same for the tag groups being considered. What are the wild releases from Bonneville-McNary? What stocks are they from and are they truly natural-spawning salmon? I note that many of the Columbia River chinook stocks have been diluted, if not replaced, with hatchery salmon at some point.

c) validity of inference and conclusions reached

The evidence for differential ocean survival rates for lower Columbia and Snake River chinook salmon stocks is circumstantial and quite weak. The results indicate different recovery locations between the lower Columbia and Snake River CWT hatchery chinook. However, the river reach recovery area interaction was not significant when bootstrapped.

A comparison is made between hatchery and wild fish released from Bonneville-McNary, but not between hatchery and wild Snake River chinook, presumably because few wild chinook salmon have been tagged. This is yet another example in PATH of trying to infer the survival of the wild Snake River stocks of interest from hatchery-reared smolts.

It is suggested on page 2 that the cumulative ocean survival rate may depend on the age of return. It would be fairly easy to test whether the entire life-cycle survival from spawner to recruit depends on age of return. I don't think this is the case, but have never explicitly tested for it.

The recovery locations do not map the entire migration pathway of the fish. A salmon could migrate to Alaska and be off caught off Oregon. Obviously, it is very difficult to measure age-specific mortality rates of salmon at sea. A working hypothesis, that is supported by available data (e.g. Healey, 1991, Pacific Salmon Life Histories), is that survival is size dependent. Thus we would expect mortality to be greatest during the first year of ocean residence and to decrease thereafter. Finally, there is not direct evidence of greater survival of chinook salmon migrating to Alaska versus those remaining in the California current.

The hypothesis being advanced here is that different salmon stocks may experience different ocean conditions. The implication for the PATH analyses is that potential differences in ocean survival could be confounded with dam passage mortality. Because of this potential confounding, it is not possible to include all these factors as parameters in the linear models. Therefore we must resort to Occam's razor as suggested in point 6 on page 3 and decide which is the most likely explanation of the observed survival rates. The "Chapter 5" analyses built a compelling case that the differences in survival rates between up and down river stocks correspond with the numbers and timing of dam construction. Furthermore, the passage mortality could be predicted from water transit time, which is known to affect in-river survival.

- d) suggestions for improvements and extensions to the analytical approaches used

Given the sparseness of the recoveries a presence-absence model might fit better. There may also be a two-step procedure for dealing with the large number of zeros in the data. The first step fits presence-absence and the second step fits the numbers within the presences. The reference on ZIP regression is:

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34:1-14.

Outlier deletion does not make statistical/biological sense here. The "outliers" are the non-zero recoveries, so it doesn't make sense to delete them. Bootstrapping makes much more sense for sparse data such as these.

- e) opportunities for integration of the different component analyses into an adaptive management approach
- f) relative priorities for future work on these analyses

This is a high priority task but it will always be limited by the sparseness of CWT returns. Given the low numbers of wild Snake River chinook salmon relative to hatchery smolts and the closure of ocean fisheries on the west coast, further tagging experiments are probably not feasible.

**Reviewer:** Brian Dennis

**Title of Paper:** *Update on Ocean Distributions of Coded Wire Tagged Spring/Summer Chinook*

**Authors:** C. Paulsen and T. Fisher

Comments:

a) scientific soundness of the methodology

The methodology is difficult to evaluate, because it is insufficiently explained. The results are the paper is written as an internal memorandum, quoting other internal memoranda, with outside readers left scratching their heads. I can list a few examples of confusing statements (below), but I must emphasize that such examples are numerous. The authors (and PATH authors in general) must start writing with outside consumers in mind.

(p. 6) Clarify "the value of the parameter is constrained to equal 1". What, exactly, is equal to 1,  $R_{ifj}$  or  $\ln(R_{ifj})$ ? The footnote #1 is not clear.

(p. 7) Write the probability model and the random variable explicitly. I assume it is a \*scaled\* Poisson distribution, not a Poisson distribution.

(p. 7) What does "bootstrapped the model" mean? What is the "sample space" from which 500 draws were made?

b) general suitability of the data for use in the analyses

I think the data are suitable, but it is hard to tell. The data description (p. 4) was written for team investigators with many joint workshops under their belts. Resist the temptation to send every reader running to the Beamesderfer et al. tome. Give readers a brief but self-contained description of the data used.

c) validity of the inferences and conclusions reached

I must defer this question until I understand the analyses better.

d) suggestions for improvements and extensions to the analytical approaches used.

A main question seems to revolve around the problem of excess zeros. Apparently, if I understand correctly, there are zero counts in the data over and above what is predicted by the (scaled) Poisson distribution. They might need to develop a mixed model approach: a logit model to account for the presence/absence of zeros, and then a conditional Poisson model to account for recoveries given fish are present. This type of model would require some programming, and a Skalski-type statistician, to accomplish.

e) opportunities for integration of the different component analyses into an adaptive management approach

None yet. However, if the conclusions of this study hold up, system managers will have to place a high priority on long-term, stage-based, capture-recapture studies in order to separate the ocean and watershed influences on salmon stocks.

f) relative priorities for future work on these analyses

I think it is vital that this work be developed into a scientific paper, completely describing the data and methods for outside readers. The conclusions are highly important to the overall PATH analyses.

**Reviewer:** Saul B. Saila

**Title of Paper:** *Update on Ocean Distribution of Coded Wire Tagged Spring/Summer Chinook*

**Author:** C. Paulsen and T. Fisher

Comments:

a) scientific soundness of the methodology

This reviewer has no specific comments on the scientific soundness of the methodology used. However, comments on the applications and assumptions involved with generalized linear models have been made previously.

b) general suitability of the data for use in the analyses

Not much can be said about the suitability of data, other than that all that was available seemed to have been utilized after suitable screening.

c) validity of inferences and conclusions reached

The conclusions appear to be valid under the assumption that the model requirements were met.

d) suggestions for improvements and extensions to the analytical approaches used

Suggestions for improvement: This review suggests a specific method to account for zeros in the data. The method suggested is the so-called "delta" method. The delta method is based on Aitchison (1955) and assumes that the raw data follow the delta distribution (Aitchison and Brown 1957). This assumption simply requires the non-zero values in the data to be approximately log-normally distributed.

The delta estimates closely follow the normal distribution for sample sizes  $N =$  or  $> 15$  (Owen and DeRuen 1980), and the standard error estimate can be used to construct confidence intervals for the mean. Pennington (1983 and 1986) demonstrated the use of the delta distribution with trawl survey data.

References:

Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. J. Amer. Statistical Assoc. 50:901-908.

Aitchison, J. and J.A.C. Brown. 1957. The log-normal distribution. Cambridge University Press, Cambridge, Massachusetts.

Owen, W.J. and T.A. DeRuen. 1980. Estimation of the mean for log-normal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants. Biometrics 36:707-719.

Pennington, M. 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39:281-286.

Pennington, M. 1986. Some statistical techniques for estimating abundance indices from trawl surveys. Fisheries Bulletin 84(3):519-525.

Outlier detection comments: The presence of outliers may be an indication of natural variability, weaknesses in a model, the data or both. Measurement errors, judgment errors, execution faults, computational errors, or a pathological case among sound data can lead to extreme values—so-called outliers. A general rule this reviewer is familiar with says that if there are at least ten individual values, then a value may be discarded as an outlier provided it lies outside the range  $\bar{0} \pm 4s$ , where the mean  $\bar{0}$  and  $(s)$  the standard deviation are computed without the suspected value, which is the outlier.

- e) opportunities for integration of the different component analyses into an adaptive management approach

There seems to be opportunities for integration of some of these results into an adaptive management approach.

- f) relative priorities for future work on these analyses

Moderately high priority for future work on these analyses is suggested—especially work with the delta distribution

### **Section 3**

## **Prospective Analyses of Spring/Summer Chinook Reviewed in FY 1997**

**with comments by the Scientific Review Panel**

**Reviewer:** Jeremy Collie

**Title of Paper:** *Applying Decision Analysis to PATH: Discussion Paper*

**Author:** C. Peters and D. Marmorek

Comments:

a) scientific soundness of the methodology

On pages 4 and 11 there are questions about how to combine different effects. A reasonable null hypothesis is that stage-specific survivals are multiplicative. Therefore we should expect actions on different life stages to be additive on a log scale. There is no mystery here; this is basic population theory.

Where distributions of parameters are available, I think it is preferable to use the distribution itself, rather than discretizing the distribution with some probability level. For example there is a distribution of year effects (delta) that can be used.

b) general suitability of the data for use in the analyses

c) validity of inference and conclusions reached

d) suggestions for improvements and extensions of the analytical approaches used

I recommend that a model-based approach to decision analysis be used as much as possible. A mathematical model can organize relevant information about management actions, performance measures, key hypotheses and uncertainties into a systematic framework that is more powerful than lists and flowcharts. I realize that not all the relevant information can be quantified. On the other hand the influences of the "four Hs" have been, or are being, quantified in functional forms that can be incorporated into life-cycle models; climatic effects can be incorporated as autocorrelated random variables.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

It is not clear to me that any decision will result from this decision analysis. My candid reaction to this paper was "get on with it."

**Reviewer:** Brian Dennis

**Title of Paper:** *Applying Decision Analysis to PATH: Discussion Paper*

(comments also apply to subsequent papers: "Tasks for PATH decision analysis..." by Peters, Marmorek, and Jones, and "Alternative structures for a hydro action decision tree" by Peters, Marmorek, and Deriso)

**Authors:** C. Peters and D. Marmorek

Comments:

a) scientific soundness of the methodology

Methodology is Bayesian, and therefore it is unsound. See my 1996 Ecol. Applications paper and my review of the paper by Rick Deriso (pkg #1). I'm obviously completely out of step with this direction that the PATH team appears to be going in, and I don't have anything constructively useful to say about how to improve Bayesian analysis.

Yes, decision analysis is a highly useful way of organizing complex information and possible outcomes. However, the fundamental scientific error can be found among the "eight essential elements" (p. 3) of decision analysis. The "essential element" #4, probabilities of uncertain states of nature, is the wolf in sheep's clothing which forces the Bayesian approach. "Uncertain states of nature" do not have probabilities. Rather, uncertain decision-makers and stakeholders have \*beliefs\*. The probabilities of hypotheses do not exist as concrete quantities in nature; rather, they exist only in people's minds. Therefore, there is no "rigorous method for determining these probabilities" (p. 13).

b) general suitability of the data for use in the analyses

c) validity of the inferences and conclusions reached

d) suggestions for improvements and extensions to the analytical approaches used.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses



**Reviewer:** Saul B. Saila

**Title of Paper:** *Applying Decision Analysis to PATH: Discussion Paper*

**Author:** C. Peters and D. Marmorek

Comments:

a) scientific soundness of the methodology

I strongly believe that a formal decision analysis approach to PATH is a potentially useful procedure. To those individuals who would like a relatively painless but useful introduction to this subject area. I recommend the book by R.T. Clemens (1991) entitled *Making Hard Decisions: An Introduction to Decision Analysis*, PWS-Kent Publishing Co., Boston, Massachusetts.

b) general suitability of the data for use in the analyses

The authors have already commented on the problem of adequate data for use in a formal decision analysis. I cannot add more to this point.

c) validity of inferences and conclusions reached

There were no specific inferences made, and the general conclusion is that decision analysis seems an appropriate approach for PATH-related problems.

d) suggestions for improvements and extensions to the analytical approaches used

I don't have any specific suggestions except to state that there is a lot of software and approaches that can be applied to decision making. I would like to indicate some of the software that I have used to a modest extent. They include the following:

1) Fuzzy Judgment Maker

Fuzzy Logic Inc.  
1160 Via Espana  
La Jolla, CA 93037

2) HIPRA3+ (**H**ierarchical **P**reference **A**nalysis)

(based on the analytical hierarchy process)  
Helsinki University of Technology  
System Analysis Laboratory  
02150 Espoo, Finland

3) Insight

Microsoft Excel add on which includes:

decision trees  
Sam L. Savage is the author of this software.

I have also written and modified some software for decision analysis.

e) opportunities for integration of the different component analyses into an adaptive management approach

I believe there are numerous opportunities for integration of components into an adaptive management decision approach.

- f) relative priorities for future work on these analyses

High priority should be given for future work. Part of this priority time should be spent in applying various types of software to gain experience and to settle on one or more specific programs.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Alternative Structures for a Hydro Action Decision Tree*

**Author:** C. Peters, D. Marmorek and R. Deriso

Comments: Perhaps he didn't intend it for general distribution, but I enjoyed reading Randall Peterman's candid letter.

a) scientific soundness of the methodology

I reiterate the need for a generic hierarchical life-cycle model as opposed to competing model formulations (page 4).

I have very little confidence in the ability of detailed mechanistic models to predict survival rates after drawdown. As pointed out in Randall Peterman's note, the existing passage models are overparameterized and were tuned to existing data as opposed to being statistically fit. As a result, only weak validity tests of the passage mortality models are possible.

With respect to Approach A of estimating probabilities, there is a limit to the complexity of hypothesis chains that can be tested with spawner-recruit data. With many stage-specific factors included in the model and data existing only for the entire life-cycle, there is an increasing risk of attributing the variations in survival to the wrong factor. There appears to be a concern that this may have already happened in the Chapter 5 retrospective analyses (cf. Paper 2-2). A logical conclusion is that the complexity of prospective models should be no greater than can be supported with existing stage-specific data.

b) general suitability of the data for use in the analyses

c) validity of inference and conclusions reached

d) suggestions for improvements and extensions to the analytical approaches used

With respect to delayed mortality, it seems as if there should be a hypothesis that delayed mortality is greater for transported than non-transported fish. Otherwise it is difficult to understand why the return rates have been so low for transported fish.

I lean toward more aggregated prospective models of passage mortality. For example, Deriso et al. fit a model (#36) in which passage mortality was proportional to water transit time (WTT), and it was one of the "top models." CRISP/FLUSH could be used to estimate the change in WTT associated with natural river drawdown and the resulting WTT could be used in a prospective version of Model 36. This relatively simple model could handle both the management actions that are listed in Table 1: flow augmentation and drawdown.

Peterman stated the need to subject each model (hypothesis) to strong tests, preferably field experiments. Along these lines, I suggest that a valuable hydro management action would be the drawdown of 1 or 2 Lower Snake dams, instead of 4. It is possible that this more limited experiment would be sufficient to meet survival objectives. More importantly, the drawdown of even a single Snake River dam could substantially reduce uncertainty about implementation delays, sediment releases, flow patterns, and predation rates. Such an experiment would provide a strong test of the passage mortality models.

- e) opportunities for integration of the different component analyses into an adaptive management approach

Decision analysis, as defined in the PATH documents, seems to be the same thing as adaptive management as defined in Walters' 1986 book. The only practical difference I see is that, with adaptive management, there is expected to be feedback between management decisions and the probabilities assigned to alternative hypotheses. For the Columbia River decision analysis, there is no such feedback, and for this reason, I find it difficult to see how the component analyses can be integrated into a truly adaptive approach.

- f) relative priorities for future work on these analyses

The hydro action decision tree is obviously a cornerstone of PATH. However, I agree with Randall's conclusion that we should temper our expectations about the ability of the decision analysis to rank management actions.

**Reviewer:** Brian Dennis

**Title of Paper:** *Alternative Structures for a Hydro Action Decision Tree*

**Author:** C. Peters, D. Marmorek and R. Deriso

See reviewer's comments for the paper "*Applying Decision Analysis to PATH: Discussion Paper*".

**Reviewer:** Saul B. Saila

**Title of Paper:** *Alternative Structures for a Hydro Action Decision Tree*

**Author:** C. Peters, D. Marmorek and R. Deriso

Comments:

a) scientific soundness of the methodology

Since the methodologies are still in a developmental stage, their scientific soundness remains to be evaluated. However, a few suggestions are offered for consideration. I believe that a fuzzy logic approach to addressing ambiguity and uncertainty in decision analyses might be a reasonable alternative. See the next item concerning suitability of this data for more specifics.

b) general suitability of the data for use in the analyses

I'm convinced that the data to be used for the analysis is the best available. It is my opinion that Approach C (to use expert judgments to assign probabilities) is particularly very useful. However, I believe that the assignment of possibilities using fuzzy logic may also be a useful alternative. Prof. Lotfi Zadeh, the inventor of fuzzy logic, is at the University of California, Berkeley, California. I think his advice on the appropriateness of fuzzy logic to elements of the decision model would be extremely helpful. At the University of British Columbia, there is also Prof. Clarence de Sylva, Department of Mechanical Engineering, who has published a book entitled *Intelligent Control, Fuzzy Logic Applications*. He may also be helpful and is even closer to the people at ESSA. It is my opinion that fuzzy rule-based models have been applied with considerable success to many complex and non-linear systems, including reservoir management at a sophisticated level and medical diagnosis, as well as to control physical systems.

c) validity of inferences and conclusions reached

Comments on the validity of inferences and conclusions are premature and should be made upon completion of the work.

d) suggestions for improvements and extensions to the analytical approaches used

Suggestions for improvements and extensions have already been made in a specific context.

e) opportunities for integration of the different component analyses into an adaptive management approach

Opportunities for integration into an adaptive management approach are deferred.

f) relative priorities for future work on these analyses

I continue to believe the relative priorities for future work and acceleration of the present level should be very high, and this should include advice from outside experts.

## References

- Eddy, D.M., Y. Hasselblad, and R. Shachter. 1992. Meta-analysis by the confidence profile method. Academic Press, New York.
- Jamir, T.V., A. Huyer, W. Pearcy, and J. Fisher. 1994. The influence of environmental factors on the oceans survival of Oregon hatchery coho (*Oncorhynchus kisutch*) (unpublished report). College of Oceanic and Atmospheric Science, Oregon State University, Corvallis, Oregon. pp. 115-138.
- Pearcy, W.G., J.P. Fisher, G. Anne, and T. Meguro. 1996. Species associations of epipelagic nekton of the North Pacific Ocean 1978-1993. Fisheries Oceanography 5(1):1-20.
- Pech, N. and F. Laloe. 1997. Use of principal component analysis with instrumental variables (PCAIV) to analyze fisheries catches data. ICES Journal of Marine Science 54:32-47.
- Scott, E.L. 1979. Correlation and suggestions of causality sality: Spurious correlation. pp. 239-251. In: Multivariate Methods in Ecological Work. L. Orleci, C.R. Rao, and W.M. Steteler (eds.). International Cooperative Publishers House, Fairland, Maryland.
- Walker, H.A. and S.B. Saila 1984. Incorporating climatic variation and hydrographic information into shrimp yield forecasts using seasonal climatic component models. pp. 57-99. In: Proceedings of the Shrimp Yield Prediction Workshop. A.M. Landry and E.F. Klima (eds.). TAMU-SG-110, April 1996.
- Warwick, R.M. and K.R. Clarke. 1993. Comparing the severity of disturbances: A meta-analysis of marine macrobenthic community data. Marine Ecology Progress Series 92:221-231.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Prospective Analysis of Spring Chinook of the Snake River Basin*

**Author:** R. Deriso et al.

Comments: This paper is a good first step at prospective analysis.

a) scientific soundness of the methodology

I am not familiar enough with the Bayesian model or the MCMC algorithm to critique their use here. This approach does seem to be on the cutting edge of fish population dynamics. A more conventional Monte Carlo prospective simulation model could also be used. The important point here is that the BSM does incorporate the main sources of variability and uncertainty.

The compensatory model seems to be the one I suggested in an earlier review, though parameterized slightly differently. I am not sure why you need the additional compensation term on page 5. Couldn't you just fix  $p$  at some positive value to achieve the same effect?

b) general suitability of the data for use in the analyses

c) validity of inference and conclusions reached

d) suggestions for improvements and extensions of the analytical approaches used

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

I think that this paper makes a useful contribution because it shows how much passage mortality would need to be reduced to meet stated survival and recovery thresholds. Also, because it is the first prospective model we have seen, I give it high priority for future work.



**Reviewer:** Brian Dennis

**Title of Paper:** *Prospective Analysis of Spring Chinook of the Snake River Basin*

**Authors:** R. Deriso et al.

Comments:

a) scientific soundness of the methodology

I do not subscribe to Bayesian approaches. My criticisms of Bayesian methods are on the record (Dennis 1996, *Ecological Applications* 6:1095-1103), and there is not much point in repeating them here. My comments have to date not been refuted or even challenged. Since these comments, a devastating critique of Bayesian methods by a philosopher of science (and accomplished mathematician) has appeared (Deborah Mayo, 1996, *Error and the growth of experimental knowledge*, Univ. Chicago Press). I urge the scientists involved in PATH to tackle the issues of Bayesianism directly before building a major portion of their analyses on this cornerstone. Is Bayesianism really going to give anything meaningful to PATH, except trouble?

Bayesians, like creationists, use slick language to substitute for scientific knowledge. Consider the following quote from the manuscript:

"The Bayesian approach allows for the calculation of the probability distribution for alternative hypotheses about Chinook population dynamics by admitting uncertainty about the fundamental parameters governing our model of their dynamics."

The quote speaks about THE probability distribution as if it is some fundamental constant of nature. In fact, hypotheses don't have probability distributions; hypotheses are either right or wrong (and much of the time we don't know which). Instead, different stakeholders have different BELIEFS about which hypotheses are right or wrong and so there are actually many subjective probability distributions about hypotheses floating around. The economic stakes involved in some scientific questions (i.e. do forest management practices affect salmon) are so high, that misrepresentation by stakeholders is rewarded (it PAYS to be wrong). You can bet that the timber or energy companies could redo the analyses proposed in this manuscript and come up with entirely different results.

The quote speaks of "admitting uncertainty about the fundamental parameters..." as if other approaches in science do not admit uncertainty. In fact, quantifying such uncertainty is a fundamental goal of other approaches. Bayesianism, however, deals with this uncertainty dishonestly, by substituting beliefs for evidence. (Note that "noninformative" priors are actually highly belief-laden; the term "noninformative" used by Bayesians is Orwellian doublespeak. (See Dennis 1996).

I propose this: Get the Bayesian prospective analysis all set up and ready to run. Let Washington WaterPower provide the priors. Is this agreeable to PATH?

NO? Of course not. But PATH scientists' beliefs do not belong in the analysis any more than Rush Limbaugh's do. One person's knowledge is another person's ignorance. This is why good science seeks to eliminate the people from the equation and let the data do the talking.

Bayesian approaches are antiscientific. Expert witnesses in the courts will cut the BAYESIAN PATH results to pieces. I urge the scientists involved in PATH to "admit uncertainty" and proceed with filling gaps in our knowledge in an honest, reliable fashion.

The OJ Simpson verdict is a classic case of Bayesian decision making.  $H_0$ : he didn't do it.  $H_1$ : he did do it. No matter what the evidence says, the priors decide.

- b) general suitability of the data for use in the analyses
- c) validity of the inferences and conclusions reached
- d) suggestions for improvements and extensions to the analytical approaches used.
- e) opportunities for integration of the different component analyses into an adaptive management approach
- f) relative priorities for future work on these analyses

I recommend that PATH abandon the Bayesian approach to prospective analysis, and concentrate on more conventional PVA approaches.

**Reviewer:** Saul B. Saila

**Title of Paper:** *Prospective Analysis of Spring Chinook of the Snake River Basin*

**Author:** R. Deriso et al.

Comments:

a) scientific soundness of the methodology

I don't consider myself adequately qualified to provide an objective analysis of this material. However, I tend to agree with a recent paper by Dennis (1996) which describes some of the difficulties with Bayesian reasoning. My limited knowledge indicates that the tools used for Bayesian inference include posterior distribution, credibility intervals, and Bayes factors in contrast to confidence intervals and P values.

b) general suitability of the data for use in the analyses

I have commented in a previous review on the general suitability of the data, and I do not have any compelling reason to comment further.

c) validity of inferences and conclusions reached

I cannot critically evaluate the validity of the inferences and conclusions. However, I tend to have a strong personal bias against any model predictions or projections over time frames extending to several decades. It is my personal belief that environmental conditions existing after such long periods will almost certainly be very different and will significantly affect results from models which do not incorporate the dynamics of environmental and anthropogenic changes.

d) suggestions for improvements and extensions to the analytical approaches used

I cannot offer any specific suggestions for improvements or extensions.

e) opportunities for integration of the different component analyses into an adaptive management approach

I cannot comment on opportunities for integration of this material into an adaptive management framework.

f) relative priorities for future work on these analyses

No opinion on relative priorities.

**Reviewer:** Jeremy Collie

**Title of Paper:** *Depensation, Performance Standards, and Probabilities of Extinction for Columbia River Spring/Summer Chinook Salmon*

**Author:** L. Botsford

Comments: The treatment of depensation in prospective analyses presents a dilemma. On one hand there is little statistical evidence for depensatory recruitment rates at low stock sizes (e.g. Myers et al. 1995). In this Columbia River situation, evidence for depensation is weak even though many of the stocks have, at some time, been reduced to very low numbers of spawners. The author demonstrates that depensation could be obscured by aging and stock misclassification errors. On the other hand, biological theory and some experience suggest that population processes will change at very low abundance levels, increasing the risk of stock collapse. There are examples of collapsed stocks that have failed to recover even when fishing has been stopped. The northern cod stock, for example, has shown little sign of recovery even though commercial fishing was halted in 1992.

a) scientific soundness of the methodology

What happened to Table 1?

Note that there is a formal way of removing the common year effects from the stock-recruitment data without assuming a stock-recruitment model. The linear model is:

The least-squares solution for the year effects ( $\tau$ 's) is given by the calculations for two-way ANOVA.

b) general suitability of the data for use in the analyses

c) validity of inference and conclusions reached

The comparison of PATH survival to quasi-extinction (page 14) is unclear because the text and figures are not well explained. It seems as if the PATH survival criteria differ slightly from the risk of quasi-extinction because the former is defined as the proportion of time spent above the threshold ( $T_2$ ). More standard measures of risk in population biology are the probability of falling below  $T_2$  at any time during the simulation, and the probability of being below  $T_2$  at the end of the time period (e.g. 24 years). I think this definition of survival would be more exactly equivalent to quasi-extinction.

d) suggestions for improvements and extensions to the analytical approaches used

The standard Ricker-type models that are being used for prospective analyses do not include a change in dynamics at low population levels. On the contrary, the Ricker model predicts higher per capita recruitment at low abundance. I can think of three ways to deal with this dilemma

1. Stick with the standard Ricker stock-recruitment model, and use the probability of falling below  $T_2$  as a measure of risk, even if the consequences of that risk are not quantified. This is essentially the approach taken in defining overfishing in U.S. fishery management plans. The overfishing definitions are somewhat arbitrary in that there is no clear evidence for reduction in per capita recruitment if these levels are exceeded. However, the overfishing definitions have the force of law and have promoted stock rebuilding.

2. Constrain the Ricker stock-recruitment predictions at low stock sizes. For example, ensure that simulated  $R/S$  ratios never exceed the largest observed value or some upper percentile of the observed values.

3. Fit a hypothetical depensatory stock-recruitment model (e.g. by constraining the depensation parameter in Deriso's model) and treat this as an alternative state of nature in prospective analyses. The obvious challenge here is to assign a probability to this hypothesis since there is little empirical evidence from the stock-recruitment data.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

Depensation and the risk of quasi-extinction should be incorporated into prospective analyses in some fashion. The use of threshold numbers of spawners will be adequate provided that falling below the threshold results in corrective management action within one chinook salmon generation (5 years).

**Reviewer:** Brian Dennis

**Title of Paper:** *Depensation, Performance Standards, and Probabilities of Extinction for Columbia River Spring/Summer Chinook Salmon*

**Authors:** L. Botsford

Comments:

a) scientific soundness of the methodology

This is an outstanding cautionary paper. As an expert for IDFG in the IDFG, et al. vs NMFS, et al. suit, I argued that too little attention had been paid to the possibility of biological crises in the salmon stocks at low densities. Judge Marsh quoted this argument in his "extinction vortex" opinion. This lack of attention to low density biology persists, and Botsford develops a case that low density effects not only might exist but might also seriously bias estimates of population jeopardy.

The paper is in its early draft stages, and is incomplete and sketchy at points. However, it goes about making its points clearly, and with a minimum of insider information that is bogging down some of the PATH single-spaced "gov'mint report" -style documents. A few comments on specific portions of the paper follow:

(p. 11) The "common multiplier" trend removal is unclear. Description needs to be reworded clearly and carefully; it is crucial.

(p. 12) The nature of the simulations behind Fig. 6 is unclear.

(p. 13) The wording of jeopardy concepts needs to be tightened up, with specific reference to the definitions of these concepts in stochastic processes. The "probability of being below T2" for instance, is unclear (is it the probability of attaining T2 from initial condition  $x > T2$  within  $t$  yr? Or is it the average proportion of time spent below T2 in  $t$  yr?)

(p. 15) Top paragraph unclear.

(p. 18) Middle paragraph is right on the money. There are many different ways to define jeopardy in stochastic processes.

(p. 21) Reference list is incomplete.

(p. 23-24) Missing from my copy; no Table 1, for instance.

(p. 34) Figure 7 (and corresponding text description) is quite unclear. Caption is uninformative.

b) general suitability of the data for use in the analyses

The data are suited to the analyses.

c) validity of the inferences and conclusions reached

The case developed in this paper is not conclusive, merely suggestive. However, the potential implications for jeopardy assessment of salmon stocks are far-reaching.

d) suggestions for improvements and extensions to the analytical approaches used.

There might be potential for modification and refinement of the Myers et al. (1995) meta-analyses, but targeted at just the stocks of interest to PATH, and using the stock-recruitment frameworks already developed by PATH.

e) opportunities for integration of the different component analyses into an adaptive management approach

Again, managers are warned not to do anything radical until the question of how much depensation there is in the stocks is settled. No amount of Bayesian prospective analyses will sidestep this question.

f) relative priorities for future work on these analyses

The depensation modeling work should be continued, and high priority should be given to biological studies of the potential for depensation in these stocks. Also, the persistence and recovery criteria developed by the Biological Requirements Work Group must be revised.

**Reviewer:** Saul B. Saila

**Title of Paper:** *Depensation, Performance Standards, and Probabilities of Extinction for Columbia River Spring/Summer Chinook Salmon*

**Author:** L. Botsford

Comments:

a) scientific soundness of the methodology

This reviewer believes the methodology is scientifically sound. It is suggested that these results be carefully examined, and efforts should be made to obtain more conservative PATH perspective results by the PATH participants.

b) general suitability of the data for use in the analyses

The data seems to be suitable (albeit limited) for use in the analyses.

c) validity of inferences and conclusions reached

The inferences and conclusions reached seem very valid and should serve as a warning that some of the results of PATH analyses are overly optimistic.

d) suggestions for improvements and extensions to the analytical approaches used

This reviewer believes that there is a very real need for incorporating this material into an adaptive management framework. It is also believed that a thorough analysis and evaluation of uncertainty estimates in the models from trends in the projections of abundance should be conducted.

e) opportunities for integration of the different component analyses into an adaptive management approach

It is hoped that this information can be integrated into an adaptive management framework. It seems very relevant, and it emphasizes a need for more conservative inferences in prospective PATH analyses.

f) relative priorities for future work on these analyses

This reviewer believes the priorities for future work and reconciliation with some PATH prospective analyses should be high.



**Reviewer:** Jeremy Collie

**Title of Paper:** *A Generalized Prospective Modeling Framework*

**Author:** C. Peters and D. Marmorek

Comments: This is the type of generalized, hierarchical modeling framework that I had hoped to see. It appears that, through the modeling workshop and follow-up memos, a consensus-modeling framework is emerging among the PATH analysts. I have specific comments on some of the alternative models as they are identified in Table 2 and defined in Appendix A.

a) scientific soundness of the methodology

The per-dam mortality (X) parameter includes both the direct and delayed mortality caused by Columbia River and Snake River dams.

Model 5 incorporates region-specific year effects. The likelihood ratio between models 4 and 5 should give a clear test of whether the extra \*'s of model 5 are statistically significant. If not Model 5 could be dropped from the set.

It is important to include a model with depensation at low stock sizes, even if the strength of depensation has to be fixed at some reasonable level. I don't understand the specification of this model. Where did the  $\ln(\Xi)$  come from? It seems as if the modification for additional depensation needs an extra S in the denominator to maintain the units: i.e.  $R/S = R/S(S/S_{min})^d$ . As I said earlier, it should be possible to specify depensation with one additional parameter, especially since this parameter is probably not estimable from the data.

In Models 7 and 8, the mortality terms need subscripts to show which stocks/regions they apply to. The definition of  $m$  in (7) seems circular in that (7) reduces back to Model 1. The  $m$  term does seem to be an estimate of delayed mortality, but I wonder if it is positive in all years?

The climate model (10) is not useful for prospective modeling. Even as a retrospective model, it would be overparameterized if more than a few (e.g. 3) climate variables were included. In prospective modeling, the climate variables would have to be predicted each year as random deviates with an autocorrelation structure. Therefore one might as well use a simpler description of the climatic effects to start with.

b) general suitability of the data for use in the analyses

Recall that the recruitment estimates in earlier years are not observations but have been reconstructed with averaged age structures. The spawner-recruit data may be suitable for testing Level-2 hypotheses, but not Level-3. Given that there are almost certain to be observation errors it is futile to try and partition the recruitment variability into multiple effects and at multiple spatial scales. The original Chapter 5 models were pushing the envelope of estimability and anything more complicated is bound to be over the top.

c) validity of inference and conclusions reached

I object to the extra (DEAD) mortality term ( $D_{y,j}$ ) in Equation 4-1. The inclusion of environmental effects in this term is slightly confusing and potentially misleading. Mortality terms should always be positive (since they are subtracted in 4-1), whereas environmental effects could be positive or negative and are usually specified as anomalies from a long-term mean. In the Alpha model (9), delayed mortality is

included in the  $\forall$  term which is defined to have mean zero. Deriso's memo (7-9-97) shows the equivalency of the models; the estimated productivity of the Alpha model ( $a'$ ) is basically decremented by the average delayed mortality ( $m$ ). If the  $\forall$ 's were projected with mean zero, in prospective modeling, the Alpha model could underestimate the productivity of the Snake River chinook stocks.

There seems to be a deliberate attempt to blur the distinction between anthropogenic and environmental sources of variability. In the climate model, the number of mainstem dams encountered is considered an environmental factor. One might argue that the hydroelectric operations interact with the natural environmental factors resulting in the observed recruitment variations. For example, rainfall in a given year combined with the amount of water spilled from the reservoirs, generates a flow pattern that affects the arrival time of smolts at the estuary. In this scenario, one can't isolate the effect of the dams, because it is conditional on the amount of rainfall. Models that combine anthropogenic and environmental parameters into single parameters violate the third principle stated on page 1 and should not be included in the model set. If the set of models that does not confound anthropogenic and environmental effects is the null set, the modeling exercises should be halted.

d) suggestions for improvements and extensions to the analytical approaches used

I would like to see a clear distinction between the sources of variability in the generalized stock-recruitment model:

$$\ln(R_{y,i}) = (1 - p) \ln(S_{y,i}) - a_i - b_i S_{y,i} - M_{y,i} + \varepsilon_{y,i}$$

The first three terms on the right hand side (RHS) are intrinsic stock-specific factors. The mortality term ( $M_{y,i}$ ) includes all additional anthropogenic mortality, except harvesting which was already accounted for in the run reconstructions. The mortality term can be partitioned into direct passage mortality, delayed mortality, habitat, hatchery effects, etc. The error term ( $\varepsilon_{y,i}$ ) can be partitioned to include shared year effects or environmental covariates. The columns headed Hydro-related and Environmental in Table 2 should be separate. The environmental effects are not extra sources of mortality: they are deviations (plus or minus) from the average survival.

e) opportunities for integration of the different component analyses into an adaptive management approach

I recognize that hypothesis generation is an ongoing process, but to the extent that it is data driven, hypothesis generation should largely have been completed in the retrospective analyses. If additional information exists to test hypotheses, it should already have been used in retrospective analyses to determine probabilities of the alternative models. It is important not to lose sight of the Final Report of the retrospective analyses, in which a large number of hypotheses were considered, and relative likelihoods assigned to them.

Page 7. The prospective modeling group has decided to focus for now on only two hypotheses; it is not stated which model numbers these correspond with. The two hypotheses are not fundamentally different: they just differ in parameterization. For the reasons stated above, I strongly object to parameterizations that confuse anthropogenic and environmental effects.

f) relative priorities for future work on these analyses

Of the models listed in Table 2, I think that 1,2,3,4, and 6 should go forward in prospective modeling studies. To this list I would add the model from Chapter 5 that expressed mortality as a function of water transit time (Model Number 36 in Table 5-4). This was one of the top models and relates hydro-induced mortality to a variable that is known to be important to smolt survival. All these alternative models could easily be incorporated into the prospective modeling framework that has already been developed by Deriso. The simulation framework need not be Bayesian, but I think that this list of models sufficiently captures the important uncertainties in the stock-recruitment data.

As far as I can tell, none of the other candidate models has yet completed parameter estimation, let alone been coded for prospective modeling. In the interests of time, I urge you to proceed with the prospective framework that is already in place. Otherwise, you may study these chinook salmon stocks to death.

**Reviewer:** Brian Dennis

**Title of Paper:** *A Generalized Prospective Modeling Framework*

**Authors:** C. Peters and D. Marmorek

Comments:

a) scientific soundness of the methodology

This memo discusses the modeling framework that is emerging from the PATH process and the plans for subsequent prospective analyses. According to the memo, a model for the decision analysis should be able to forecast the outcomes of alternative management actions, as well as accommodate multiple hypotheses about the system. The model should conform to four principles (listed on p. 1). The four principles are sound, and are apparently satisfied by a stochastic model of stock and recruitment that contains parameterization families corresponding to different system hypotheses.

The generalized Ricker modeling framework that is emerging from the PATH process is outstanding and state-of-the-art. A description of it, along the lines of the memo, should be written up soon into a paper for the science literature (I recommend *Ecological Applications*). Last fall, I was the reviewer for a paper that will appear soon in *Oecologia*; the title is something like "Use of covariates in density dependence analysis." The paper features weather variables in a Ricker time series model: the topic is in the wind!

The Ricker modeling framework, furthermore, is highly suited to the tasks that are being tackled by PATH. The Ricker stock-recruitment model was modified by PATH to handle the year-class problem as well as categorical and quantitative covariates (hypothesized). It potentially can accommodate regional covariation or synchrony (Dennis et al., *Ecology*, in press), which might be important to include in the prospective simulations.

The differences between AIC and BIC in model selection (p. 4) are of interest when one shifts from model evaluation to prediction, as PATH is doing in its shift from retrospective to prospective analysis. In numerous theoretical and simulation studies, the BIC is turning out to have somewhat better properties, \*if\* the goal is to pick the model that is "closest" to the true stochastic mechanism that generated the data. However, the AIC is emerging as a somewhat better tool for picking a model, \*if\* the goal is prediction (AIC is asymptotically equivalent to cross-validation). Thus, one can expect differences between the two criteria, and one can interpret these differences and use them to advantage.

I will comment mainly on the plans for the future detailed in Section 6 of the memo (p. 7-9).

1. "Agree on a generalized modeling framework (completed)." The coalescence of the PATH group around generalized stock/recruitment, and not around detailed life history simulation modeling, is very interesting from an outside observer's standpoint. Would-be modelers of the Columbia salmon have been told repeatedly by fishery scientists that the system is too complex to be addressed by simple models. But a "simple" model, artfully managed and modified by PATH, has allowed analyses of complexities and interactions that could never be addressed with a "complex" model.

2. "Generate hypotheses (ongoing)." This is an excellent step, particularly if the hypotheses are stated as explicitly as possible in the Ricker framework. A near complete listing of a priori hypotheses will strengthen the inferences, both with regard to pure hypothesis testing (picking from two models) and model selection (picking from among many models). Each family of hypotheses (ocean, dams, habitat, etc.) will have to be carefully discussed in reference to existing literature.

3. "Clearly define the initial set of alternative hypotheses to include in preliminary decision analysis framework... ." Not clear how this differs from 2. Does 2 include the completion of the retrospective analyses, so that such filtering of hypotheses implied in 3 can occur? In which case, how are 3 different from 5 below?

4. "Run the initial hypotheses through the generalized modeling framework... ." Clarify what, exactly, that means. Does it mean the following:

Stochastic simulations of all index stocks, under all hypothesized  
parameter values and noises structures, yielding prediction  
distributions for all performance measures.

Such simulations should account for estimation error in parameter values and regional covariance of stocks (e.g. Snake River), if such covariance proves important.

5. "Follow-up analyses... ." This portion is not clear to me. In particular, a lot of what is described sounds like retrospective analyses (implied in 2 above) rather than prospective analyses. How, for instance, can one do "coarse filtering" (5a, p. 8) of hypotheses based on future predictions instead of based on existing data, without waiting for the future outcomes in the system? How, also, can one use likelihoods (hypothesized models) in "assigning probabilities" (5b, p. 9) to hypotheses?

And this brings up the topic of Bayesianism again.

In my reading of PATH documents, I interpret "retrospective analysis" to mean estimation, hypothesis testing, and model selection, and "prospective analysis" to mean prediction (in support of decision making). In the Bayesian approach, prior probabilities are assigned to parameter values (hypotheses) and mixed with the likelihood to form posterior probabilities for parameters (hypotheses) in retrospective analysis. In prospective analysis, probabilities are assigned to parameters (hypotheses) and mixed with the likelihood to form prediction probability distributions for future states of the system (and their utilities). Both retrospective and prospective analyses have the same inputs. In fact, there is not much distinction between the two analyses in the Bayesian approach; both are considered types of decision making.

In other words, the Bayesian approach requires the prior probabilities in the retrospective analysis. The generalized Ricker model is the likelihood, and every parameter value or hypothetical model structure must be assigned a prior probability by the PATH team. After the data are entered into the likelihood, the fundamental structure for all Bayesian decisions, whether science or policy oriented, is complete. If it is logical to do prospective analyses with Bayesian methods, then it is ultra-logical to do retrospective analyses with Bayesian methods. One can't have it both ways.

Needless to say, I favor the conventional frequentists approach that the PATH team has adopted in their retrospective analyses. The analyses, when suitably published, will represent a landmark of fishery science, in which stock/recruitment theory and large amounts of data were combined to produce major insights into a scientifically complex (and politically charged) system. There is a sense in which the team's credibility would be damaged in the scientific and political arenas if the team's prior beliefs were entered into the retrospective hypothesis testing and model selection work. Why is the team treating forecasting so differently?

(responses to b-d below subsumed in response to a above)

b) general suitability of the data for use in the analyses

c) validity of the inferences and conclusions reached

- d) suggestions for improvements and extensions to the analytical approaches used.
- e) opportunities for integration of the different component analyses into an adaptive management approach
- f) relative priorities for future work on these analyses

**Reviewer:** Saul B. Salla

**Title of Paper:** *A Generalized Prospective Modeling Framework*

**Author:** C. Peters and D. Marmorek

Comments:

a) scientific soundness of the methodology

It is my opinion that the models under consideration from the PATH analyses describe a relatively complex large-scale system which consists of a number of sub-systems. These models attempt to emulate real world phenomena. However, they are simplifications of reality, and they contain a degree of imprecision. Modeling the system is believed to involve a trade-off between simplicity and precision. I strongly believe that including appropriate safety factors as a hedge against unsuspected model errors is an especially important consideration in situations involving public policy and common property resources, such as those in the PATH assessments. If variables are added or multiplied together in a probabilistic assessment have correlations that are either ignored or assumed to be zero, the resulting probability distribution will be underestimated in the tails. The significance of such an underestimate in prospective analyses is that the analysis will yield an erroneous and potentially dangerous miscalculation of the chance of a very high consequence event—such as extirpation of a stock. I am not confident that the methodologies to be applied will adequately account for extreme-event probabilities (i.e., the tails) of distributions resulting from analyses, which depend on dependencies among the variables involved in the assessments.

b) general suitability of the data for use in the analysis

I have repeatedly expressed concern about the quality of data for use in the analyses. I repeat this concern for the prospective modeling exercises.

c) validity of inferences and conclusions reached

The validity of inferences and conclusions will depend upon model structure, assumptions regarding dependencies, among variables, and the effective treatment of the tails of distributions for extreme-event predictions.

d) suggestions for improvements and extensions to the analytical approaches used

I can only suggest some references that deal with uncertainty in models and large-scale systems, as possible sources for further information. These include:

V. Perincherry, S. Kikuchi, and Y. Nemamotsu. 1994. Uncertainties in the analysis of large-scale systems. pp. 73-85 In: *Uncertainty Modeling and Analysis: Theory and Applications*. B.M. Ayyub and K.M. Gupta (eds.). Elsevier, New York.

Shlyakhter, A.I. 1994. Uncertainty estimates in scientific models: Lessons from trends in physical measurements, population, and energy projections. pp. 477-496. Same source as above.

e) opportunities for integration of the different component analyses into an adaptive management approach

I believe the opportunities exist, but the details remain to be resolved.

f) relative priorities for future work on these analyses

A relatively high priority should be assigned to this work.



**Reviewer:** Jeremy Collie

**Title of Paper:** *Prospective Analysis for the Alpha Model*

**Author:** J. Anderson and R. Hinrichsen

Comments: This chapter is a rough draft that probably should not have been sent out for review. The alpha model adds little to previous model formulations; it is not at all clear that the additional model parameters (the components of  $\nabla$ ) are estimable. As the alpha model apparently has not been fit to retrospective data, it is certainly not ready for prospective modeling.

a) scientific soundness of the methodology

Strictly speaking  $-M_{y,i}$  is the log of juvenile passage survival (page 1); otherwise the sign is reversed. The assumption that the  $\nabla$ 's sum to zero seems to make the definition of  $\nabla$  (16) so complicated that it would be difficult to explain in non-technical terms to decision makers (Peters and Marmorek 4th principle). The Level 3 hypotheses (Table 1) seem to be the same for transported and non-transported fish. Can the mortality of these two groups of fish be estimated and predicted separately?

To isolate the  $\nabla$  term (Eq. 30) from the stock-recruitment equation, you first need to estimate all the parameters of Eq. 1, including the  $\nabla$ 's and  $\gamma$ 's. Are you suggesting a two-stage approach, whereby the  $\nabla$ 's are first estimated in Eq. 1 and the resulting  $\nabla$  estimates are then partitioned into component sources of variation? If so, some of the variation that is being explained in the second stage may be contained in the  $\gamma$ 's from the first stage.

b) general suitability of the data for use in the analyses

Can delayed mortality (8) be estimated independently from existing data? I recall from Chapter 6 of the Retrospective Analysis that there were no direct estimates of delayed mortality and that only the ratio of delayed transport mortality to in-river delayed mortality could be estimated. Mundy et al. (1994) pointed out that there are no true controls, against which to measure delayed mortality.

c) validity of inference and conclusions reached

The authors state on page 10 that "An important result from this analysis is that the Deriso model does not express a pure common climate effect." I don't see that this result has been demonstrated.

It is true that the major variables in Table 5 have changed in the past 39 years. However I am not sure that all these changes have been coincident. To the extent that they have not been coincident, it should be possible to factor out their independent effects.

d) suggestions for improvements and extensions to the analytical approaches used

e) opportunities for integration of the different component analyses into an adaptive management approach

Equation 24 gives a suggested characterization of environmental relationships. What is the ecological rationale for this equation? Can it be validated with data? How will the climate variables be projected into the future? The description at the top of page 18 is unclear and seems to be missing some text.

f) relative priorities for future work on these analyses

I think the biggest contribution of this work is in incorporating specific hypotheses about the hydrosystem into an aggregate life-cycle model. However, until it is demonstrated that these Level-3 hypotheses are actually testable, incorporating them into prospective models should be a low priority. At any rate it is not necessary or productive to have competing, parallel model formulations. All the hypotheses that are discussed in this paper could be incorporated into the existing prospective modeling framework, and probably with simpler notation.

**Reviewer:** Brian Dennis

**Title of Paper:** *Prospective Analysis for the Alpha Model*

**Authors:** J. Anderson and R. Hinrichsen

Comments:

a) scientific soundness of the methodology

This paper was a frustrating read. On the one hand, is the best PATH paper in the current collection to consult for an explanation of the delayed mortality problem and its parameterizations? On the other hand, the paper was not enough; I still don't understand the issues. I think that the main problem is that the paper tries to do too much. I am a statistical ecologist involved mostly in terrestrial and laboratory systems, and this particular paper assumes far more background information than I have at my fingertips.

There were some statistical peculiarities that I couldn't follow:

1. Phrasing concepts in terms of standard statistical theory of linear models wherever possible would help. For instance, is the difference between the alpha and delta models that the delta model includes a year effect, while the alpha model includes a year by region interaction? Expressing the different linear models and parameterizations in matrix form (for instance,  $E(\ln R) = X1 * \beta1$  for the alpha model, and  $E(\ln R) = X2 * \beta2$  for the delta model, where  $\ln R$  is the vector of recruitments,  $X1$  and  $X2$  are full column rank covariate (design) matrices, and  $\beta1$  and  $\beta2$  are parameter vectors) might be used to advantage, to show how the \*estimates\* of  $\beta1$  and  $\beta2$  will differ.

2. (p. 4) "Taking the average of both sides..." This is not clear. What is being averaged? How do the noise terms disappear (unless the authors are actually taking \*expectations\* of both sides, which is different). The rest of p. 4 is a complete mystery.

3. (p. 6-7) I do not follow which subscripts are being summed over.

4. (throughout) Categorical variables as subscripts is really confusing.

5. (p. 9-10) I do not understand these hypotheses in relation to the alpha model. What, for instance, is  $X_n$ ? Where is it defined, and where is it in the model?

6. (p. 17) "The hypotheses A and B relating to the hydrosystem... based on the uncertainties of the regression coefficients developed in the retrospective analysis." This is unclear. Are the sampling distributions of parameter estimates going to be used as prior distributions, or what? (I think that in general the PATH team has to get a better handle on what they are actually going to do in the prospective analysis.)

7. (p. 17) "...correlation structure of the prospective analyses..." Unclear. Which random variables are being referred to here? There are many possible correlation structures to be investigated, including cross-correlations and autocorrelations in the noise terms.

No further comments on b-f below, other than to urge the PATH team to produce coherent papers for outside readers explaining (1) the delayed mortality problem and (2) the actual plans for prospective analyses.

b) general suitability of the data for use in the analyses

c) validity of the inferences and conclusions reached

d) suggestions for improvements and extensions to the analytical approaches used.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

**Reviewer:** Saul B. Saila

**Title of Paper:** *Prospective Analysis for the Alpha Model*

**Author:** J. Anderson and R. Hinrichsen

Comments:

a) scientific soundness of the methodology

I believe that the overall scientific methodology is sound. However, as stated on page 18, paragraph 2 of Final Comments, "The primary hypothesis (model) is used to estimate stock and recruitment numbers from redd counts, which are the primary data. I do not believe that the errors from redd counts are adequately considered propagated in the models.

b) general suitability of the data for use in the analyses

From the above, it is evident that I am concerned about the stock-recruit data derived from redd counts. These have apparently been accepted without testing, and I believe these data are limiting for any data based on them.

c) validity of inferences and conclusions reached

The validity of inferences from the models is clearly influenced by the quality of the primary input data mentioned above.

d) suggestions for improvements and extensions to the analytical approaches used

No suggestions for improved analytical approaches are made.

e) opportunities for integration of the different component analyses into an adaptive management approach

I have reservations about integration of these component analyses into an adaptive management framework. See the review by C. Walters (1994) to obtain a clear definition of an adaptive management approach and its implications.

f) relative priorities for future work on these analyses

The priorities for future work on these analyses depend on the quality of the inputs. In my judgment, they are not high.

**Reviewer:** Jeremy Collie

**Title of Paper:** *General Framework for Prospective Modeling with one Proposed Hypothesis on Delayed Mortality*

**Author:** P. Wilson, E. Weber, C. Petrosky and H. Schaller

Comments:

a) scientific soundness of the methodology

I was unable to rederive some of the algebra in this paper. Part of the problem might be that we were sent a photocopy of a FAX and some of the subscripts were difficult to read. How was Equation 9 derived? It implies that

$$\Delta m = P_{b,n} \lambda_n + \sum_{j=1}^5 P_{b,j} \lambda_j$$

yet  $\Delta m$  was defined as minus the log of this quantity in Equation (4). Check the algebra.

The idea of calculating ratios of annual future system survival to a base system survival (page 8) has some merit, as long as we are interested in the relative improvement in survival over the base case. However, the survival standard is expressed in absolute number of spawners and therefore the prospective model must also predict absolute numbers. Why is the term  $\ln(\omega_p/\omega_r)$  repeated in Equation 16?

Please provide a more complete explanation of Figure 1. It looks interesting but I don't understand the variables that are being plotted.

b) general suitability of the data for use in the analyses

c) validity of inference and conclusions reached

As in the paper by Peters and Marmorek, delayed mortality is expressed as the difference between the MLE estimate of total mortality and the direct estimate of passage mortality (page 3  $\Delta m = m - M$ ). This definition seems somewhat circular; given an estimate of the total passage mortality, why not just use that? What if  $m < M$ ?

Can delayed mortality ( $\lambda$ ) be estimated independently from existing data? I recall from Chapter 6 of the Retrospective Analysis that there were no direct estimates of delayed mortality and that only the ratio of delayed transport mortality to in-river delayed mortality could be estimated. I don't recall seeing any evidence that the total survival of transported and non-transported smolts is systematically different. A working null hypothesis is that the total survival of transported smolts is no higher than in-river smolts. Until this hypothesis is falsified, it may not be necessary to distinguish between transported and non-transported smolts in prospective models.

I agree with the rationale for including a shared year effect in the environmental component of the model (page 10). As noted by the authors, there is a strong environmental signal in the northeast Pacific, and its effects are shared by stocks of different species from different regions. Therefore it is not unreasonable to expect a shared year effect for Columbia River spring-spawning chinook salmon.

- d) suggestions for improvements and extensions to the analytical approaches used
- e) opportunities for integration of the different component analyses into an adaptive management approach
- f) relative priorities for future work on these analyses

Again, the biggest contribution of this work is in incorporating specific hypotheses about the hydrosystem into an aggregate life-cycle model. However, until it is demonstrated that specific hypotheses about delayed mortality are actually testable, incorporating them into prospective models should be a lower priority than running prospective models that can be parameterized with existing data.

**Reviewer:** Brian Dennis

**Title of Paper:** *General Framework for Prospective Modeling with one Proposed Hypothesis on Delayed Mortality*

**Authors:** P. Wilson, E. Weber, C. Petrosky and H. Schaller

Comments:

a) scientific soundness of the methodology

Overall, this memo is way over my head, and I do not follow its technicalities. It is a memo for insiders about some differences in model parameterizations. As a general reader not steeped in the scientific problems, I would need an explanation of these numerous quantities and processes.

The notation appears peculiar, excessive, and needs to be redesigned with readers in mind. For instance, as far as I can tell, the subscripts used are both categorical and numerical; thus "N<sub>xy</sub>" means different things depending on the context of the subscripts x and y.

From my perspective, the "delayed mortality" question is being addressed in by PATH in disjointed, disorganized forms. I see memos and appendices, but little direct analyses, results, and conclusions. The analyses are of course in the preliminary planning stages. However, it seems to me that delayed mortality is a hypothesis of importance to the system, and its evaluation should be properly a part of retrospective analysis instead of prospective analysis.

I have no particular comments related to b-f below, except to say that the scientific question of delayed mortality is important and high priority should be given to clarifying and executing the analytical approaches that will be used to quantify it.

b) general suitability of the data for use in the analyses

c) validity of the inferences and conclusions reached

d) suggestions for improvements and extensions to the analytical approaches used.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

**Reviewer:** Saul B. Saila

**Title of Paper:** *General Framework for Prospective Modeling with one Proposed Hypothesis on Delayed Mortality*

**Author:** P. Wilson, E. Weber, C. Petrosky and H. Schaller

Comments:

- a) scientific soundness of the methodology

The methodology proposed seems to be scientifically sound.

- b) general suitability of the data for use in the analyses

Again, the major concern is the suitability of the data. I do not think that the data available justify the elegant expressions derived for testing various hypotheses concerning mortality

- c) validity of inferences and conclusions reached

The validity of the inferences is a function of the quality of the data. I tend to be pessimistic about the outcomes of the exercise.

- d) suggestions for improvements and extensions to the analytical approaches used

I have no suggestions other than improved data, and this cannot be manufactured.

- e) opportunities for integration of the different component analyses into an adaptive management approach

I have no comments in this matter.

- f) relative priorities for future work on these analyses

I am not inclined to encourage future work under the circumstances.



**Reviewer:** Jeremy Collie

**Title of Paper:** *Integration of PATH Chapter 6 Survival Goals with Chapter 5 Passage Mortality Estimation Procedures*

**Author:** C. Toole

Comments: This paper makes a valuable contribution in relating passage mortality estimates to interim survival goals. While I agree with the rationale, the loose notation makes it difficult to follow the calculations. In several cases I found myself redoing the calculations to identify what the numbers represented. For example,  $M$  generally refers to an instantaneous rate of natural mortality; in this paper it is used to represent a proportional mortality. A more formal mathematical notation would help to make the calculations clearer.

a) scientific soundness of the methodology

I think that Tables 3 and 4 could be explained better. As stated on page 8, Table 3 calculates what combinations of juvenile survival, adult survival and environmental conditions correspond with an aggregate survival that matches brood years 1957-1965. The explanation of  $e^*$  is what confused me. Where did the 0.28 come from? Table 3 indicates how much juvenile survival would need to be increased to maintain overall survival the same as for the 1957-65 brood years. Equation 4 partitions this increase equally between direct and delayed passage survival.

The rationale of Section 3 is that smolt-to-adult return rate (SAR) is a component of spawner-to-adult survival, but that it has been measured for only a subset of the brood years. Can the parameters of the spawner-recruit model be used to reconstruct SAR for the missing brood years? I agree with the rationale of this section but had difficulty following the mathematical notation until I recreated it. I think this is a valid approach and have given an alternative derivation below that follows more directly from population modeling.

The stock-recruitment model has aggregate parameters that potentially all contain components of SAR. The  $a-bS + \gamma$  term contains natural mortality including the SAR stage. The  $-X_n$  term all applies to the SAR stage, except for survival through the uppermost reservoir. The shared year effects (\*) potentially apply to the entire life-cycle, but to the extent that they are ocean effects, they may apply mostly to the SAR stage.

p.14. Note that the effect of "fallback" on adult survival rates would seem to depend on whether the salmon were falling back at the upstream or downstream dam.

p.18. From the equations (see below) one would expect an exponential relationship between SAR and \*.

b) general suitability of the data for use in the analyses

The reason that the estimate of "non-passage S" is so high in 1972 is that it had the second lowest estimate of passage S (Table 5). By accounting for all the additional sources of mortality, you get more smolts than were there to start with. When you multiply and divide four terms, all of which are measured with error, it is not surprising that there are some outliers. However, I am not sure that the grounds for rejecting BY 1972 (p. 18) are sufficient. Non-passage mortality of the 1971 brood year was anomalously low because of low estimated SAR (Table 5). This low-high pattern in adjacent years could result from aging errors. To objectively reject or downweight a year, you would need to know something about how the data were collected that year (e.g. sample sizes, assumptions made, etc.).

c) validity of inference and conclusions reached

As defined by Deriso (1996) the X-type mortality term does include both direct and delayed passage mortality. Therefore delayed mortality for the downstream stocks would be included in this term of the model. If the non-hydro year effects are believed to be region specific, they should be modeled explicitly. Deriso did fit models of this form (Model Numbers 22-28 in Table 5-4 of the Retrospective Analysis) and they can be compared with the corresponding models with a common year effect. "The models which include regional year effects generally performed worse than the empirical models which exclude this alternative hypothesis." Including regional year effects did not appear to change the : mortality, which suggests that the year effects were not confounded with passage mortality. In short, it is desirable to have a 1:1 correspondence between hypotheses and models.

d) suggestions for improvements and extensions to the analytical approaches used

To derive an estimator of smolt-to-adult survival, start with the extended Ricker stock-recruitment model:

$$(1) \quad R_{MO} = Sm \cdot e^{a_2 - \beta_2 Sm + \varepsilon_2} e^{-Xn - \mu} e^{M_1}$$

where  $R_{MO}$  is the aggregate return of Snake River chinook salmon to the mouth of the Columbia River and  $Sm$  is the aggregate number of smolts measured at the uppermost dam. The subscript 2 is used to identify this as the 2nd stage of the life cycle; thus  $a_2$ ,  $\beta_2$  and  $\gamma_2$  apply only to the smolt-to-adult stage.  $M_1$  is the mortality in the uppermost reservoir, which must be added back in because smolts are enumerated at the downstream end of the reservoir.

Since SAR is calculated based on returns to Lower Granite Dam, the returns must also be adjusted for in-river mortality. Recruits at Lower Granite Dam ( $R_{LG}$ ) are

$$(2) \quad R_{LG} = R_{MO} N ACR$$

where  $N$  is 1-Harvest and  $ACR$  is the Adult Conversion Rate. Then

$$(3) \quad SAR = \frac{R_{LG}}{Sm} = e^{a_2 - \beta_2 Sm + \varepsilon_2} e^{-Xn - \mu} e^{M_1} \phi \cdot ACR$$

The terms of (3) can then be rearranged to isolate the unknown Ricker parameters. I call the aggregate term Non-Passage Survival (NPS) in keeping with the last column of Table 5.

$$(4) \quad NPS = \frac{SAR}{e^{-Xn - \mu} e^{M_1} \phi \cdot ACR} = e^{a_2 - \beta_2 Sm + \varepsilon_2}$$

NPS is smolt-to-adult survival that would be expected without the additional sources of mortality and is therefore a useful baseline for evaluating hydropower impacts. Equation 4 can also be written

$$(5) \quad \ln(NPS) = a_2 - \beta_2 Sm + \gamma_2$$

I took the aggregate smolt numbers from Table 6 of Petrosky and Schaller (1996) for brood years 1962-82 (see attached spreadsheet). The estimate of  $\beta_2$  is positive (i.e. a negative slope) as expected but it is not

significantly different from 0 (see spreadsheet) and had a very minor effect on  $\ln(NPS)$ . Thus we cannot reject the null hypothesis of density-independent smolt-to-adult survival, at least for the range of smolt numbers that have been measured. This analysis should be redone when the SAR's for brood years 1990-93 become available. The low smolt numbers from these brood years will provide a stronger test of density dependence. If density-dependent smolt survival occurred, we would need to measure smolt numbers to predict SAR. In we accept the null hypothesis of no density-dependence at the smolt stage we can safely ignore the  $\exists_2 Sm$  term in (5). This result is also interesting in that it attributes virtually all of the density dependence in recruitment to the freshwater stage (see Petrosky and Schaller 1996).

The second question is how strongly the error term  $\gamma_2$  is related to  $*$ ? The regression equation becomes

$$(6) \quad \ln(NPS) = a_2 + c *$$

The important point is that the expected relationship between NPS and  $*$  is exponential not linear as plotted in Figure 2 of Toole. The relationship described by (6) is positive as expected (see attached figure) with 1971 and 1972 as notable outliers. There are at least two reasons for lack of colinearity between these variables.

1. The error term  $\gamma_2$  contains additional variability not contained in the common year  $*$ . If so, using  $*$  to predict SAR would underestimate the true variability in SAR.
2. As both variables are measured with error, a functional regression might be more appropriate.

Finally we have three estimates of the expected NPS:

- |   |       |
|---|-------|
| 1. Geometric mean of the NPS estimates: | 0.202 |
| 2. $\exp(a_2)$ from Equation (5):       | 0.229 |
| 3. $\exp(a_2)$ from Equation (6):       | 0.208 |

Given that the regression models were not significant, the geometric mean is probably the most reliable. At any rate, non-passage survival estimates of around 20% give a baseline against which to compare the 2-6% SAR goal.

For prospective models, SAR can be projected with

$$(7) \quad SAR = e^{a_2 + \varepsilon_2} e^{-Xn - \mu} e^{M_1} \phi \cdot ACR$$

- e) opportunities for integration of the different component analyses into an adaptive management approach

ASSUMPTION 1B on page 12 essentially says that the Snake River stocks experienced differential ocean survival after the 1969 brood year and that this mortality coincided with hydropower development. This is the time-treatment interaction described by Walters, Collie and Webb (1989), which requires a staircase experimental design to resolve. To the extent that dam construction was staggered over time, it is possible to resolve the incremental mortality associated with each dam. The staircase design could be run in reverse by sequentially decommissioning dams and measuring the change in survival rate. The decommissioning would have to be carefully timed so as not to be in phase with a ~20-year climate cycle. This is about the most informative and consequential adaptive management experiment that could be devised for the Snake River chinook salmon.

f) relative priorities for future work on these analyses

I give high priority on relating mortality estimation procedures to the interim survival goals. It will help immensely to have direct mathematical equivalencies between the two. The partitioning of life-cycle survival into stage-specific components deserves more attention because it helps identify where impacts are occurring.

**Reviewer:** Brian Dennis

**Title of Paper:** *Integration of PATH Chapter 6 Interim Survival Goals with Chapter 5 Passage Mortality Estimation Procedures*

**Authors:** C. Toole

Comments:

a) scientific soundness of the methodology

This is another internal report written for PATH investigators. To me, the final 2/3 of the paper was not very readable, as I do not have the mental recall of all the assumptions, hypotheses, data bases, jargon, and analyses that PATH has been so intensely involved in these past two years (What was in Chapter 5? What was in Chapter 6?) I just can't follow these arguments in any reasonable time frame.

A background paper for outsiders on delayed mortality is needed. This paper started out well, and I found Figure 2 (p. 21) to be the start of something really useful. I had tried to draw a picture like that for several of the papers in the current batch, but I just didn't have the necessary understanding. I wish someone would write a paper about Figure 1 to lead people like me by the hand through the various complexities.

b) general suitability of the data for use in the analyses

c) validity of the inferences and conclusions reached

d) suggestions for improvements and extensions to the analytical approaches used.

e) opportunities for integration of the different component analyses into an adaptive management approach

f) relative priorities for future work on these analyses

**Reviewer:** Saul B. Saila

**Title of Paper:** *Integration of PATH Chapter 6 Interim Survival Goals with Chapter 5 Passage Mortality Estimation Procedure*

**Author:** C. Toole

Comments:

a) scientific soundness of the methodology

The scientific soundness of the methodology in Parts 1 and 2 of this chapter is not questioned. However, it is my opinion that the section, which suggests a method for representing smolt-to-adult, return rate (SAR) in prospective modeling analyses does not appear to be scientifically sound enough. In general, any regression which explains less than 50 percent of the variability of the observations is suspect. In this case,  $r^2 = 0.27$ , and this is considered inadequate for prediction purposes.

b) general suitability of the data for use in the analyses

The same criticism holds regarding the suitability of data. If the initial estimates of mortality are subject to significant error, then any mathematical manipulation of these estimates will increase the error. Partitioning mortality from initial estimates, which have substantial error, can lead to misguided inferences from hypothesis tests related to these estimates.

c) validity of inferences and conclusions reached

The validity of the inference is again related to the quality of the inputs.

d) suggestions for improvements and extensions to the analytical approaches used

I suggest that the analytical approaches used may be more sophisticated than the data justify.

e) opportunities for integration of the different component analyses into an adaptive management approach

No comment on integration.

f) relative priorities for future work on these analyses

The relative priorities for future work on these analyses should be low on the basis of available data.